مجلة الروائز
ISSN: 2602-5914
المجلد:04 العدد:01 (السنة: 2020) ص ص: [05-34]

معايير الاختبار الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية في المدارس الحكومية في محافظة طولكرم.

Standards of Good Testing from the Point of View of Primary and Secondary School Teachers in Public Schools in Tulkarm Governorate.

زياد بركات*
جامعة القدس المفتوحة، فلسطين
zeiadb@yahoo.com

تاريخ الإرسال: 2020/05/19 **تاريخ القبول: 2020/06/30**

ملخص:

هدفت الدراسة الحالية إلى تحديد مدى الالتزام بمعايير الاختبار الجيد لدى معلمي المرحلتين الأساسية والثانوية بمحافظة طولكرم في فلسطين. لهذا الغرض طبقت استبانة مكونة من (36) فقرة لقياس مدى التزام المعلمين بهذه المعايير على عينة مكونة من (170) معلماً ومعلمة، تم اختيارهم بطريقة عشوائية طبقية وفق متغيري الجنس والمرحلة التعليمية. وتكشف النتائج عن مستوى مرتفع لالتزام المعلمين بمعايير بناء الاختبار الجيد، وكانت المعايير العشرة الأكثر التزاماً لديهم على الترتيب الآتي: درجة وضوح الأسئلة وسهولة لغة الاختبار، وطريقة توزيع الدرجات على أسئلة الاختبار مناسبة، والتنوع في صعوبة الأسئلة لمراعاة الفروق الفردية، وأسئلة الاختبار متوافقة مع ما تم شرحه، وتطابق الاختبارات مع وصف ومفردات المقرر، ويشتمل الاختبار على أسئلة تميز الطلبة المتفوقين، ويغطي الاختبار المستويات المعرفية المختلفة، ويكتب الاختبار بصيغة تناسب المرحلة الأكاديمية للطالب، وتطابق الاختبار مع الوقت المخصص، وتوزيع أسئلة الاختبار على المادة العلمية يحقق شمولية الاختبار. ومن جهة أخرى، أظهرت النتائج وجود فروق دالة إحصائياً في مستوى التزام المعلمين بمعايير الاختبار الجيد وفق مغير الجنس لصالح المعلمات، ووفق متغير المرحلة التعلمية لصالح معلمي المرحلة الثانوية. بينما عدم وجود فروق دالة إحصائياً في هذا المستوى وفق متغيرات: التخصص، والمؤهل العلمي، وسنوات الخبرة.

-الكلمات المفتاحية : الالتزام؛ المعايير؛ الاختبار الجيد؛ معايير الاختبار الجيد؛ المرحلة الأساسية؛ المرحلة الثانوية.

Abstract:

The present study aimed to determine the extent of adherence to good testing among primary and secondary teachers in government schools in Tulkarm Governorate, Palestine. For this purpose, a questionnaire consisting of (36) items was applied to measure the extent of teachers' adherence to these standards on a sample of (170) male and female teachers, who were chosen in a stratified random manner according to the gender and educational stage variables. The results revealed a high level of teachers' commitment to the criteria for building a good test, and the ten most important criteria were in the following order:The degree of clarity of questions and the ease of the test language, the method of distributing grades to

* المؤلف المرسل

appropriate test questions, the diversity in the difficulty of the questions and taking into account individual differences, and the test questions are compatible with what was done explanation, and the conformity of the tests with the description and vocabulary of the course, the test includes questions that distinguish outstanding students, and the test covers various cognitive levels, and is written in a form appropriate for the academic stage of the learner, and the test coincides with the allotted time, and the distribution of exam questions on the scientific subject achieves the comprehensiveness of the test. On the other hand, the results showed that there are statistically significant differences in the level of teachers' commitment to the criteria of good testing according to the gender variable in favor of female teachers, and according to the variable of the educational stage in favor of the secondary stage. And the absence of statistically significant differences at this level according to variables: specialization, educational qualification, and years of experience.

Keywords: Adherence; criteria; good testing; criteria of good test; primary and secondary teachers.

مقدمة:

تشكل الاختبارات أهم أدوات القياس التي يمكن استخدامها في تقويم أداء الطلبة في المراحل المختلفة، وذلك من أجل الحكم على مستوى تحصيلهم في المقررات المختلفة، والتقويم في التعليم يمثل حجر الزاوية في العملية التعليمية، على اعتبار أنه يمثل مصدر اتخاذ القرار حول مستوى النجاح أو الفشل في تحقيق الأهداف، ولهذا فإن مصداقية التقويم توفر للبرنامج التعليمي أساسيات النجاح، لأن النجاح يتوقف على فعالية وسائل وإجراءات التقويم، ومدى اعتماد عملية التقويم على المعايير الأساسية، والمبادئ التي ينبغي أن تؤخذ بعين الاعتبار، والامتحان الجيد هو الذي يتمتع بدلالات الصدق والثبات الجيدة، وحيث أن صدق الاختبار وثباته يعتمدان على خصائص فقراته الاختبارية، لذلك كان لا بد من الاهتمام بمعرفة الشروط الواجب توافرها في الفقرات الاختبارية وفقاً لمعايير الامتحان الجيد، ونظراً للشكوى المستمرة من المشرفين عن قصور الاختبارات من نواح عدة, سواء من حيث الإعداد، أو الغموض، أو عدم قدرتها على التمييز، وعدم اشتمالها على المعايير الجيدة.

للتأكد من كفاية معلومات شخص ما أو من مقدار مهارته في أداء عمل معين فإنه يتطلب منه الإجابة على مجموعة من الأسئلة، أو القيام بعمل ما أو حل مشكلة معينة ثم يتم تقييم أداءه أو ما فعله, وبعبارة أخرى فإنه يخضع لاختبار شفوي أو عملي أو تحريري أو لجميعها معاً، أو يعطى عملاً لينجزه ثم يتم تقييم مدى جودة أدائه له (حماد, 2011). وإذا أمعن النظر في حياة الإنسان نجد أنه كان يختبر ويمتحن نفسه باستمرار، فإذا كان يعتقد أنه يستطيع القيام بعمل معين, وضع اعتقاده هذا موضع الاختبار بمحاولته القيام بذلك العمل، وقد كانت حياة الإنسان عباره عن سلسله من الاختبارات العملية, فالمهارات وأساليب السلوك المختلفة أو الأعمال التي نجح في أدائها تبناها وكررها عندما طلب منه ذلك أو عندما امتحن, وهكذا كان الإنسان يقوم بالعديد من الاختبارات اليومية كلما صنعَ أداة أو اخترع شيئاً جديداً أو حاول امتحان قدرته على أداء الجديد من الأعمال (Friedlend & Folt, 2015).

للتقويم أساليب متنوعة حيث تعتبر الاختبارات وسيلة من وسائل التقويم الرئيسة لقياس المستوى التحصيلي للطلبة, والتعرف أيضاً على مدى تحقيق المادة الرئيسية لأهدافها وكذلك معرفة مواطن القوة والضعف فيها ومعرفة مدى التقدم الذي أحرزته المدرسة، وعندئذ يمكن في ضوء ذلك تحسين العملية

التعليمية وصولاً إلى الأمثل والأفضل, فالاختبارات لها دوراً كبيراً في العملية التعليمية والتربوية حيث أنها من أهم الأدوات لجمع المعلومات اللازمة لعملية التقويم التربوي والرقي به, فالاختبارات هي عبارة عن أسلوب منظم لمقارنة أداء شخص أو مجموعة أشخاص طبقاً لمستوى معين في الأداء, بمعنى آخر هو عبارة عن أداة أو وسيلة أعدت بطريقة منظمة من مجموعة مرتبة من المثيرات لتقدير خاصية سلوكية محددة لدى الطالب للتعبير عنها في صورة كمية أو رقمية, وهناك تعريف شامل لمفهوم الاختبار هو مجموعة من المفردات "الأسئلة" التي تعطى للطالب للإجابة عنها شفوياً أو تحريرياً وقد تكون موضوعية أو مقالية أو رسوماً أو أشكالاً تستخدم للمقارنة والقياس (Anastasi & Urbina, 1997).

مشكلة الدراسة

إن طبيعة الاختبارات العامة تؤثر في وجهة نظر المعلمين وطرق تدريسهم، فمعظمهم يضطرون إلى توجيه عناية كبيرة متساوية إلى جميع تفاصيل برنامج المادة الدراسية بصرف النظر عن مدى ميل الطلبة إلى ما يقومون بدراسته. إن الاختبارات بشكلها الحالي قلبت غرض التربية والتعليم من الإعداد للحياة والمستقبل إلى الإعداد للامتحان لأن الكثير من الأنظمة في كثير من الدول العربية، ومن بينها فلسطين تشجع متعلميها على الحفظ غيباً وبهذا تقتل الأصالة والقدرة على التفكير والإبداع والاختراع.

ويمثل الاختبار عنصراً رئيساً من عناصر التعليم الصفي، إذ يمكن تضمينه كوسيلة معينة على التعليم، وذلك في أثناء العملية التعليمية، وفي نهايتها، هذا بالإضافة إلى الغرض المرئي، في كونه أداة لجمع المعلومات من أجل اتخاذ القرار المناسب، سواء على المستوى الصفي، أو على مستوى المؤسسات التربوية ، بشكل عام. وتُعد عملية التقويم الناجحة في أي مدرسة ركيزة أساسية لتحقق الأهداف التربوية والتعليمية، من خلال تنظيم أساليب تقويم التحصيل الدراسي وإجراءاته، وتحديد مستوى تحصيل الطالب، والتعرف إلى مدى تقدمه نحو تحقيق الغايات والأهداف، كذلك تزويد الطالب والقائمين على العملية التعليمية بالمعلومات اللازمة، من أجل تحسين مستوى التعلم، ورفع كفاءة المناهج وأساليب التدريس، وتطوير عمليات التقويم وإجراءاته والمراجعة المستمرة لها، وفق الأسس العلمية، فضلاً عن المساهمة في الحدِّ من مشكلات الرسوب، وما يترتب عليه.

وبناء على ذلك، فإن كل عمل أو نشاط يتطلب وقفة تأمل وتفحص لهذا العمل في كل مراحله حتى الانتهاء منه وقد يحتاج العمل في أثناء ذلك إلى تعديل وتحسين حتى ينجز كما يجب أن يكون تطبيقاً، وهذا ما يطلق عليه بناء الاختبارات وعملية تقييم هذه الاختبارات وفق معايير الاختبار الجيد، إذ لا تستقيم عملية التعلم والتعليم إلا بالتقويم والتقييم, فتعد الاختبارات المدرسية ضمن نظام التقويم المهم؛ إلا أنها الوسيلة الوحيدة المعتمدة في تقويم نتاجات التعلم، وبالتالي تقويم كل العملية التعلمية التعليمية في أطرها العامة، وبذلك فإن إشكالية الدراسة الحالية تتمحور حول السؤال الرئيس الآتي: ما مستوى الالتزام بمعايير الاختبار الجيد لدى معلمي المرحلة الأساسية والثانوية في المدارس الحكومية في محافظة طولكرم؟

أسئلة الدراسة:

سعت هذه الدراسة إلى الإجابة عن التساؤلات الآتية:

1. ما أهم معايير معايير الاختبار الجيد من وجهة نظر معلمي المرحلة الأساسية والثانوية في المدارس الحكومية في محافظة طولكرم؟

2. هل توجد فروق دالة إحصائياً في معايير الاختبار الجيد من وجهة نظر معلمي المرحلة الأساسية والثانوية في المدارس الحكومية في محافظة طولكرم تبعاً إلى متغير الجنس؟
3. هل توجد فروق دالة إحصائياً في معايير الاختبار الجيد من وجهة نظر معلمي المرحلة الأساسية والثانوية في المدارس الحكومية في محافظة طولكرم تبعاً إلى متغير المرحلة التعليمية؟

هل توجد فروق دالة إحصائياً في معايير الاختبار الجيد من وجهة نظر معلمي المرحلة الأساسية والثانوية في المدارس الحكومية في محافظة طولكرم تبعاً إلى متغيرات: التخصص العلمي، والمؤهل العلمي، وسنوات الخبرة؟

أهمية الدراسة

تبرز أهمية هذه الدراسة من الناحيتين النظرية والتطبيقية؛ فمن الناحية النظرية: ينبغي على المعلم الاهتمام بالتقويم لما له من أهمية ودور كبير في إنجاح العملية التعليمية التعلمية، وأن عليه مراعاة تنفيذ جميع أنواعه، والتنويع في أدواته وأساليبه. أما من الناحية التطبيقية فتنبع أهمية الدراسة من ما تقدمه عملية تقييم الاختبارات والتزامها بمعايير الاختبار الجيد في معرفة أداء الطالب وقدرته على التعلم في المجالات المختلفة المعرفية والانفعالية والحس حركية. وفي مدى توفر أدوات التقييم الفعالة للعمليات التي يحدث بها التعلم مثل مدى اشتمال الاختبار على الجوانب المختلفة من حيث الزمان والمكان وشمولية الأسئلة، إضافة إلى معلومات مهمة حول الاختبارات من شأنها مساعدة المعلمين وواضعي المناهج على تحسين عملية بناء الاختبارات وتطويرها، ورفع كفاية أساليب التدريس خاصة فيما يتعلق بتوضيح المهارات وتحديدها، والمعارف التي يجب أن تركز عليها الاختبارات.

أهداف الدراسة

في ضوء مشكلة البحث فإن الهدف ينصب أساساً في تحديد معايير الامتحان الجيد من وجهة نظر معلمي المرحلة الأساسية والثانوية, وتحقيق الأهداف التالية:

1. التعرف إلى أهم معايير الامتحان الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية.
2. التحقق من دلالة الفروق الإحصائية في مستوى معايير الامتحان الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية تبعاً إلى متغيرات: الجنس، والتخصص، والمؤهل العلمي، والخبرة، والمرحلة التعليمية.

الإطار النظري والدراسات السابقة

أولاً: الإطار النظري: تتمحور الدراسة الحالية حول مفهومين أساسين كالآتي:

1. **الاختبارات:** إن الاختبار عبارة عن مجموعة من الأسئلة (Questions) أو المشكلات (Problemes) صممت لتقدير المعرفة أو الذكاء أو غيرهما من القدرات والخصائص، فمفهوم الاختبار يعني ضمناً طلب الإجابة على مجموعة من الأسئلة المعدة سلفاً، بحيث يتم الحصول عليها بناءاً على إجابات المفحوص عن تلك الأسئلة، وهذه النتائج عادة ما تكون على شكل قيم عددية (Numerical Values) تعبر عن سمات وقدرات المفحوص الذي أجاب عن تلك الأسئلة (Russell, 2001). ويعرف الاختبار على أنه إجراءات منظمة تستهدف القياس الكمي أو الكيفي لمظهر واحد أو أكثر لسمة أو قدرة من القدرات عن طريق عينة من السلوك اللفظي أو غير اللفظي. ويستهدف الاختبار غرضين رئيسين

هما: الأول محاولة مقارنة نفس الفرد فى مظهرين أو أكثر بالنسبة للسمة أو القدرة المقيمة، والثاني مقارنة فردين أو أكثر فى مظهرين أو أكثر من مظاهر السمة أو القدرة المقاسة (بركات، 2018). ويوضح بيين (Bean, 1993) أن الاختبار عبارة عن مجموعة متعاقبة من المثيرات (Stimuli) نظمت لقياس بعض العمليات العقلية أو سمات الشخصية كمياً أو كيفياً، حيث يلاحظ أن هذا التعريف يكشف عن خاصيتين مهمتين للاختبار النفسي والتربوي هما: الخاصية الأولى: أن الاختبار عبارة عن تنظيم متعاقب لمثيرات، هذه المثيرات تنظم فى شكل سلسلة متتابعة من العبارات أو الجمل وفقاً لبعض المبادئ التي يتطلبها بناء الاختبار، والخاصية الثانية: أنه يمكن استخدام كل من القياسات الكمية والقياسات الكيفية بالنسبة للاختبارات النفسية والتربوية. وبذلك فإن مفهوم الاختبار يتضح في أمرين هما (Kaplan & Saccuzzo, 2009): الأمر الأول، إن الاختبار (Test) مصطلح يشير إلى مجموعة من البنود (الأسئلة أو المهام أو غيرهما) مرتبة بشكل يتيح تصحيح الإجابات أو الأداءات، حيث تستخدم الدرجات فى تقدير الفروق الفردية للمفحوصين. أما الأمر الثاني فهو أن الاختبار عبارة عن أداة قياس (Measurement Davint)تستخدم للحكم على جانب أو جوانب محددة بالنسبة للمفحوص مثل: الإنجاز أو الذكاء أو الشخصية أو غيرها.

والمتتبع لأدبيات هذا الموضوع يجد أن هناك خلطاً واضحاً بين عدة مفاهيم ذات علاقة بموضوع أدوات القياس والتقويم كالاختبارات، والمقاييس، والامتحانات، والفحوص، والاستفتاءات، والاستقصاءات، والاستبانات، والقوائم، والاستبارات، والاستخبارات، والروائز. ويستخدم الباحثون هذه المفاهيم أحياناً بشكل مترادف وأحياناً بشكل متناقض، ولتوضيح الفروق بين هذه المفاهيم للوصول إلى فهم سليم لها من أجل توظيفها توظيفاً صحيحاً في البحوث التربوية والنفسية والاجتماعية، يتم تعريف هذه المفاهيم والتفريق بينها نظرياً وإجرائياً كالآتي (بركات، 2018؛ Holland & Rubin, 2009؛ Hodges, 2009؛ Kaplan & Saccuzzo, 2009؛ Domino & Demino, 2006؛ Ebel, 2006؛ Schmit & Drasgow, 2002؛ Kelley, 2001؛ Russell, 2001؛ Anastasi & Urbina, 1997):

- الاختبار (Test): إجراء منظم لقياس عينة من السلوك أو إجراء منظم لملاحظة ووصف سمة أو أكثر من سمات الفرد. والاختبار مرتبط أكثر في قياس الأداء في المجالات المعرفية (الذكاء، والاستعدادات والتحصيل)، وهو عبارة عن مجموعة من الأسئلة أو البنود لكل منها إجابة واحدة صحيحة فقط مثل اختبارات التحصيل أو اختبارات الذكاء والقدرات العقلية وغير ذلك من الاختبارات التي تقيس مجموعة من الحقائق. ورغم تعدد وتنوع أدوات القياس التي يمكن للمعلم أن يستخدمها في جمع البيانات عن طلبته، إلا أن الاختبار هو أكثر هذه الأدوات أو الأساليب شيوعاً واستخداماً بين المعلمين في المدارس، ومن ثم تعد الاختبارات النفسية أهم أدوات القياس النفسي فهي الأدوات العلمية لدراسة التفرد المتمايز لكل شخص في تكيفه مع بيئته.

- المقياس (Scale): المقياس أداة قياس للجوانب المعرفية وكذلك الوجدانية أو تلك التي يتم قياسها على سلم يتألف من عدد من الدرجات (٣)، أو (٥) أو (٧). ومقاييس الشخصية لا يجوز وصفها بأنها اختبارات لكونها لا تطلب من المفحوص التفاعل أثناء التطبيق، ولكونها أيضاً لا تحمل بالنسبة للمفحوص معنى الامتحان، إن القياس يعد أكثر اتساعاً من الاختبار، فيمكن قياس بعض الصفات أو الخصائص باستخدام الاختبارات أو بدونها، فقد يستخدم لقياس خصائص وسمات معينة ببعض الأساليب. هذا، وتتغير

طبيعة الاختبار تبعاً لتغير طريقة تطبيقه، فعندما يطبق لأول مرة على مجموعة من الأفراد، فهو يعد اختباراً نفسياً، وعندما تتوالى مرات تطبيقه على نفس المجموعة، فإنه بذلك يتحول إلى تجربة من تجارب التعلم تقيس مدى تحسن الأداء تبعاً لعدد مرات المحاولة، وعندما تحلل أخطاء التطبيق وتتم الاستفادة من هذه الأخطاء، فإن الاختبار قد يتحول بذلك إلى مقياس من مقاييس الشخصية بعد أن كان اختباراً من اختبارات النواحي العقلية المعرفية، وقد تعددت أنواع الاختبارات النفسية والتربوية بدرجة تجعل الباحث يواجه صعوبة في حصرها، ومن ثم يتم تصنيف الاختبارات والمقاييس إلى مجموعات بناءً على أسس ومحكات مختلفة تم العرض لها في فصل سابق.

- الامتحان (Examination): وهناك أيضاً خلطاً بين مفهومي الامتحان والاختبار والفرق بينهما هو في وظيفة كلٍ منهما؛ ففي حين يتعلق الامتحان بقياس الأداء في المجالات التحصيلية، فهو يرتبط في العملية التعليمية فقط، أمّا الاختبار فيستخدم للقياس في عدة مجالات غير التحصيل الدراسي. وعرف الاختبار على أنه أداة قياس يتمّ من خلالها فحص الطالب أو المتقدم، للتأكّد من مدى إتقانه للمادّة أو المهارة التي اكتسبها ضمن خطوات وشروط، أمّا الامتحان فيعرف على أنّه: الوسيلة التي يتم فيها تقييم مدى قدرة الطالب على تحقيق الأهداف التعليمية، ومدى قدرة المعلم على إتقان تدريس المادّة التعليمية، والفرق بين الاختبار والامتحان أيضاً يتضح في الاختبار هو معيار للطالب بأنّه يمتلك الأدوات الصحيحة التي تمكنه من اجتياز المادة التعليمي، أمّا الامتحان فيتّخذ صفة الرسميّة أكثر، فيقيم مستوى الطالب، وعند اجتيازه تمكنه من الانتقال من مرحلة إلى المرحلة التي تليها من حيث المادة التعليمية، ويقيس الاختبار مدى إتقان الطالب لمفهوم أو درس واحد، أمّا الامتحان يقوم بقياس مدى إتقان الطالب لعدة مفاهيم أو وحدة دراسية، والاختبار يتمّ داخل وخارج المؤسسة التعليمية، مثل اختبار السياقة، أما الامتحان فهو خاص بالمؤسّسات التعليمية، والامتحان يرتبط بالعملية التعليمية، أمّا الاختبار فيرتبط في عدة مجالات أخرى.

- الاستبيان (الاستبانة) (Questionnaire): الاستبانة لغة كلمة مشتقة من الفعل استبان الأمر، بمعنى أوضحه وعرفه، والاستبانة بذلك هي التوضيح والتعريف لهذا الأمر. واصطلاحاً: هي نموذج مطبوع يحتوى على مجموعة من الأسئلة المرتبة بأسلوب منطقي مناسب، مقسمة حسب وحدات معينة، وتدور حول موضوع معين، يتم وضعها في استمارة ترسل لعينة من المفحوصين بالبريد، أو يجرى تسليمها باليد، أو تنشر على صفحات الجرائد والمجلات، أو من خلال وسائل التواصل الاجتماعي كالفيسبوك والتويتر وغيرها، أو تبث عن طريق التلفاز أو الإذاعة، أو على شبكة الانترنت، يجيب عليها هؤلاء المفحوصون بدون مساعدة من الباحثين سواء في فهم الأسئلة أو تسجيل الإجابات عليها، ثم يقومون بإرسالها إلى الباحث أو الهيئة المشرفة على البحث بعد الانتهاء من ملئها، ويستخدم مصطلح استمارة الاستبانة للإشارة إلى الاستبانة فيقال استمارة الاستبانة، وتعتبر هذه الاستمارة من أكثر أدوات جمع البيانات شهرة وانتشاراً في العلوم الاجتماعية والتربوية والنفسية، وهي تفيد تقريباً في كافة البحوث الاجتماعية فهي تستخدم في البحوث الكشفية لجمع أكبر قدر من البيانات عن الظاهرة موضوع الدراسة، وتستخدم أيضاً بكفاءة أكثر في البحوث الوصفية لتقرير عما توجد عليه الظاهرة في الواقع، كما تستخدم في البحوث التجريبية وغيرها.

- الاستبار (الاستبارة) (Interviewing Schedule): لا يزال بعض الباحثين يخلطون بين الاستبانة والاستبارة ويستخدموهما كمترادفين، ولهذا يتوجب بيان الفرق بينهما على النحو الآتي: عادة ما تستخدم الاستبانة في بحث كل الظواهر الاجتماعية والإنسانية في مجالاتها المختلفة دون حصر هذا الاستخدام في نوع معين من الأبحاث، وعلى وجه الخصوص في المجتمعات غير الواسعة التي لا تتعدى

مفرداتها بعض المئات، بقصد الوصول إلى بيانات ذات صلة بالدراسة، من خلال استعمال عشرات الأسئلة، التي تتناول العديد من الجوانب المختلفة بالمفحوصين. بينما ينحصر استخدام الاستبارة من حيث موضوع الأسئلة في تحقيقات الرأي ذات الصلة بقياس الآراء، وتحديد المواقف من قضايا سياسية معينة لدى جمهور الرأي العام العريض المتكون عادة من آلاف المفحوصين، عن طريق استخدام عدد قليل من الأسئلة لا يتجاوز عددها حدود الصفحة الواحدة في كل الحالات. وخلاصة القول أن الفرق بين الاستبانة والاستبارة هو في عدد الأسئلة وحجم المجتمع؛ ففي الاستبانة يكون عدد الأسئلة كبير وحجم المجتمع محدود، بينما في الاستبارة يكون عدد الأسئلة قليل وحجم المجتمع ضخم وغير محدد.

- الاستفتاء (Referendum): وهو عبارة عن مجموعة من الأسئلة أو البنود التي تدور حول موضوع واحد أو عدة مواضيع وليس لها إجابات صحيحة أو إجابات خاطئة إذ أن المطلوب هو معرفة رأي الفرد أو نوعية استجابته. أما الاستفتاء العام فهو الذي تقوم به الصحف والمؤسسات الخاصة والعامة أو الدولة لرصد اتجاهات الرأي العام بالنسبة لموضوع أو قضية اجتماعية أو اقتصادية أو سياسية معينة، ويبحث الاستفتاء العام في المواقف المختلفة لفئة من الناس تستخلص منها بعض النتائج التي تساعد في تكوين فكرة عامة عن أي موضوع أو قضية، وهناك الاستفتاء الذي قد ينص عليه دستور الدولة كوسيلة أساسية لمعرفة رأي الشعب في مسالة أو أكثر متعلقة بنظام الدولة العام أو رأيه بالموافقة أو الرفض في المسألة المعروضة عليه في الاستفتاء دون وسيط.

- القائمة (Inventory): وتسمى أحياناً ببطاقة الملاحظة وهي دليل أو قائمة من الفقرات أو العبارات تهدف إلى تقييم مدى وجود اتجاهات أو اهتمامات أو سلوكات محددة لدى مجموعة من الأفراد، وغالباً ما تعرف كأدوات -ورقة وقلم- تستخدم في تقويم الشخصية أو الاتجاهات أو الميول، وتعرف أحياناً بأدوات التقرير الذاتي (Self-Reports). ويندر استخدامها كمقاييس تحصيل أو مهارات أو قدرات أو استعدادات. وهى تحتوى على قائمة بالخطوات والفعاليات والسلوكيات التي يسجلها المعلم أو القائم بالملاحظة، وهى تمكن الباحث من ملاحظة ما إذا كان السلوك قد حدث أو لا وهى لا تسمح بأن يصنف نوعية السلوك المحدد أو درجته أو تكراره ويمكن استخدام قوائم التدقيق في معظم المجالات الدراسية والفعاليات، والتي تنطوي على إنجاز معين أو عملية أو عمليات متسلسلة مثل تشغيل المجهر ومحضر العينات وتدوين النتائج.

- الفحص (Checkup): نوع من الإجراءات الطبية التي تعمل لفحص أو تشخيص أو تقييم مرض أو قابلية أو تحديد الاستجابة للعلاج ، وذلك من خلال إجراء فحوص مخبرية أو شعائية أو غير ذلك قبل التشخيص والعلاج.

- الرائز (روائز) (Quotient): هو تقرير موضوعي لسلوك الشخص المفحوص بعيداً عن تدخل العوامل الذاتية للفاحص أو لمصحح الرائز، وعليه فإنه لا بد من تصحيح الرائز وضبط شروطه في كل مرة يستخدم فيها، سواء استخدم على المفحوص نفسه أم على مفحوصين آخرين. ويبدو أن في استخدام مصطلح الرائز دقة لا تتوافر في استخدام مصطلحات أخرى تبدو مرادفة له كالمقياس، والاختبار، والاستبانة، والقائمة، والاستخبار، لأنه يرتبط بضرورة توافر الخصائص القياسية الضرورية والمعروفة بالصدق والثبات والمعايرة والتقنين، وهي خصائص لا ضرورة لتوافرها جميعاً في المصطلحات الأخرى ولا سيما في المعايرة والتقنية.

- الاستخبار (A question): وهو الاستفهام، وهو طلب الفهم وقيل الاستخبار ما سبق أولاً ولم يفهم حق الفهم، وفي بعض الكتب الاستخبار هو طلب الخبر. ولغة هو السؤال عن الخبر، وخبر الأمر أي علمه، والاستخبار في علم النفس هو طريقة من طرق قياس الشخصية أو نوع من المقابلة المقننة، ويشمل الاستخبار على مجموعة من الأسئلة أو العبارات التقريرية التي تقدم مطبوعة غالباً، ويجيب عنها المفحوص بنفسه بالكتابة غالباً، في حدود فئات محددة مثل نعم ولا، وأوافق أو لا أوافق، أو ينطبق أو لا ينطبق علي. ويدور الاستخبار حول جوانب وجدانية انفعالية أو خاصة بالسلوك في المواقف الاجتماعية، ويجيب عنها المفحوص على أساس معرفته لمشاعره وانفعالاته وسلوكه الماضي أو الحاضر، وتقدر درجات الاستخبار ويفسر بطريقة موضوعة سلفاً.

- بطارية الاختبار (Battery): مجموعة مختارة بعناية من الاختبارات المترابطة التي تقيس سمة أو قدرة واحدة أو أكثر، يتم عرضها على مجتمع معين من الأفراد، وعادة ما يتم تقييم مستوى الأداء على أحد اختباراتها أو على عدد منها أو جميعها. ومن بطاريات الاختبار المشهورة بطارية تقييم الخريجين (Graduate Record Examination) والمعروف اختصاراً (GRE)، وبطارية الاستعدادات الفارقة (Direct Aptitude Test) والمعروف اختصاراً (DAT)، وبطارية التحصيل الدراسية (Academic Achievement Test) والمعروف اختصاراً (CAT). وهذه الاختبارات تم تقنينها على مجتمع العينة بحيث يمكن مقارنة نتائج جميع الطلبة على هذه الاختبارات. وتصمم هذه الاختبارات ثم تعطى للطلبة كوحدة واحدة بقصد تقييم الأداء في العديد من الموضوعات المتعلقة بالمحتوى (Content) والمجالات النفسية (Psychological areas)، مثل: الذكاء والشخصية، وتقدير الذات (Self-esteem). ومن أشهر بطاريات الاختبارات تلك المستخدمة في تقييم التحصيل المدرسي، وهي اختبارات تتضمن اختبارات جزئية في موضوعات مدرسية منفصلة تستخدم نتائج هذه الاختبارات مجتمعة لتعطي مؤشراً واحداً يستخدم في اتخاذ العديد من القرارات المتعلقة بالطالب أو برنامجه الدراسي.

- الاختبارات غير الرسمية (Informal Tests): وتعرف أحياناً بالاختبارات من صنع الشخص، وهي الاختبارات التي يتم بناؤها أو تطويرها من قبل معلم المدرسة أو أستاذ الجامعة أو مسؤول البرنامج التدريبي بغرض قياس مستوى تحصيل الطلبة أو كفاءة المتدربين، وغالباً ما تعرف هذه الاختبارات بالامتحانات.

- الاختبارات الرسمية (المقننة) (Standardized) (Formal Tests): وهي الاختبارات التي يتم بنائها وتطويرها من قبل فريق من المختصين ولصالح مؤسسة معينه، كما يتم تطبيقها على عينة كبيرة نسبياً من الأفراد ويتم تطوير دليل (Manual) لها يتضمن تعليمات خاصة بالتطبيق والتصحيح ومعلومات عن مؤشرات الصدق والثبات، وخصائص الفقرات وتفسير النتائج عليها باستخدام معايير (Norms) خاصة يتم اشتقاقها. وتعرف الاختبارات المقننة بالاختبارات المنشورة. ومن أمثلتها: اختبارات الذكاء وقوائم الاتجاهات ومقاييس الشخصية واختبارات التحصيل الوطنية مثل امتحان الثانوية العامة في معظم البلدان العربية.

وبذل علماء القياس جهداً كبيراً نحو توفير اختبارات تتمثل فيها جميع درجات أو مستويات الكفاية ويعبر عن نتائجها بنفس وحدات القياس وتتمتع بالصفات الأتية: تحمل معنى واحد في الاختبارات حتى يسهل مقارنة درجات المفحوصين على هذه الاختبارات، وتتألف من وحدات متساوية في جميع الحالات، ولها صفر حقيقي يعني انعدام الصفة (Angoff, 1982). على أنه في مجال تطوير الاختبارات النفسية والتربوي فقد تم التوصل تقريباً إلى حلول فاعلة بخصوص الصفتين الأوليتين، أما الصفة الثالثة فيبدو

تحقيقها صعباً جداً على أرض الواقع، بأن يصل القياس النفسي والتربوي إلى هذا المستوى من القياس بتحديد نقطة صفر حقيقية تعني عندها انعدام الصفة المقاسة (Hambleton; Merenda & Spielberger, 2005). ويتضح مما سبق أن الدرجات الخام للمفحوصين في أي اختبار لا تفيد شيئاً في تفسير هذه الدرجة لذلك ينبغي إجراء نوع من التحويل عليها لكي تصبح قابلة للمقارنة، وتحدد مكانة الفرد أو مركزه بالنسبة لأقرانه في الجماعة المرجعية، فدرجات اختبارات الذكاء والاستعدادات والشخصية والاختبارات التحصيلية المقننة وغيرها تفسر درجاتها عادة استناداً إلى جداول المعايير، والتي تكون مدونة في دليل كل اختبار ويمكن باستخدام هذه الجداول معرفة الدرجة المحولة المناظرة لدرجة خام معينة، وهذه الدرجات المحولة أو المعايير يتم إعدادها إمبيريقياً وذلك بتحديد أداء عينة من الأفراد ممثلة لمجتمع معين في الاختبار أي الجماعة المرجعية.

2. المعايير: ويقصد بالمعايير (Norms) أنها الدرجات المحولة من الدرجات الخام مثل الدرجة المعيارية، والعمرية والصفية والتائية، والمئينية، والتي تعطي الدرجة الخام معنى ودلالة، والمعايير في الاختبارات النفسية والتربوية هي أدوات قياس مقننة لعينة من السلوك بطريقة منتظمة وفقاً لقاعدة محددة، وتستخدم نتائج القياس في المقارنة بين الافراد أو المفحوصين وفقاً لمعيار أو ميزان معرف تعريفاً واضحاً مما يساعد في تفسير درجات الاختبارات (Angoff, 1971). فالدرجة الخام تأخذ معناها وقيمتها المعنوية فقط من خلال نسبتها إلى معيار محدد مسبقاً فهي على اختبار معين تكون أعلى أو أقل أو أفضل أو أسوأ من درجات أخرى على نفس الاختبار ولا تكون حسنة أو سيئة بالمطلق، وعلى هذا الأساس فإن هناك خمسة أنواع من المعايير تستخدم لهذا الغرض هي: الصفية والعمرية والمئينية والمعيارية والبروفيل الشخصي؛ والمتتبع للأدب السابق لهذا المجال يجد عدة مصطلحات ومفاهيم ذات صلة بالمعايير هي (بركات، 2018؛ Gregory, 2011؛ Schneidler, 2007؛ Flanagan, 1993؛ Anogoff, 1981):

- معايير (Standards): قائمة من التوقعات التي تصف ما نتوقع أن يتعلمه الطالب ويستطيع أداءه بعد إنتهاء فترة التعلم، في ضوء تلك المعايير يتم الحكم على الجودة النوعية لما استطاع الطالب أن يعرفه أو يؤديه، وهناك معايير متعددة، منها: معايير المحتوى (Content Standards)، ومعايير الأداء (Performance Standards)، ومعايير المناهج (Curriculum Standards)، وعموماً يمكن اعتبار المعايير وصفاً دقيقاً لمستوى التعلم المتوقع أن يحصل عليه الطالب بعد إنتهاء برنامجه الدراسي.
- معيار الأداء (Performance Standard): وصف محدد لمستوى أداء الطالب المتوقع لمهام ارتبطت بمحتوى (Content) معين، يستخدم هذا الوصف للحكم على أداء الطالب في ضوء معرفة الحد الأدنى (Cutoff- score) من الأداء المقبول لمهمة (Task) محددة، هذا المستوى من الأداء يقرره التربويون عندما يضعون الأهداف التعليمية المطلوب أن يحققها الطالب في نهاية البرنامج الدراسي. وهذه المعايير أو المحكات المقيمين (Assessors) ليحافظوا على موضوعية التقييم، كما أنها تقدم للطالب معلومات مهمة عن الأداء المتوقع منه، وتضع أمامه هدفاً يسعى لتحقيقه.
- معايير التصحيح (Scoring Criteria): مجموعة من المحكات التي يضعها المصحح مسبقاً ليحكم في ضوئها على جودة إجابة الطالب عن أي سؤال ورد في الاختبار.

- معايير المحتوى (Content Standard): معايير تضعها الإدارة التعليمية ليتم في ضوءها تحديد وتقرير ما يفترض أن يعرفه الطالب ويستطيع أدائه في كل مقرر من مقررات المنهج الدراسي عند مستوى دراسي محدد، وعندما تصاغ معايير المحتوى على المستوى الوطني تكون لها صفة العمومية، ولكنها تبقى أقل تحديداً من معايير الأداء (Performance Standards) .

- اختبار معياري المرجع (Norm-Referenced Test) اختبار يستخدم لغرض مقارنة أداء الفرد أو المجموعة (فصول، مدارس، مناطق تعليمية) بأداء مجموعة معيارية (Norm Group) تم اختيارها على المستوى الوطني، واختبارات الذكاء (I.Q) والاختبارات المقننة (Standardized Tests)هي اختبارات من هذا النوع، ويستخدم هذا النوع من الاختبارات غالباً لقياس ومقارنة أداء عينة من الطلبة بما هو متوافر من مؤشرات سائدة عن الأداء القياسي للمجتمع الطلابي - ممثلاً بعينة كبيرة (Norm- Group) على هذه الاختبارات، والهدف من هذه الاختبارات هو غالباً تصنيف الطلبة وليس لتقرير مدى تحقيقهم لبعض معايير الأداء المحددة مسبقاً.

- انحراف معياري (Standard Deviation): مقياس لمدى تشتت وتباين الدرجات حول المتوسط، وبقدر ما تتقارب الدرجات نحو المتوسط بقدر ما يتناقص الانحراف المعياري، وفي المنحني الطبيعي يتواجد ما نسبته (68%) من الدرجات في المدى المحدد بواحد انحراف معياري فوق ودون المتوسط. في اختبارات الذكاء تبين أن المتوسط هو (100)، وأن الانحراف المعياري للدرجات هو في حدود (18-15). الانحراف المعياري يساوي الجذر التربيعي الموجب للتباين (مجموع مربعات انحراف القيم عن الوسط الحسابي مقسوماً على حجم العينة (–1).

- خطأ القياس المعياري (Standard Error of Measurment): مقياس للخطأ المتوقع في درجة الاختبار، أنه المقدار الذي في حدوده تزيد أو تنقص الدرجة المشاهدة عن الدرجة الحقيقية للطالب، وكلما قل خطأ القياس المعياري كلما زادت الدقة في درجات الاختبار، وخطأ القياس المعياري هو عبارة عن الانحراف المعياري لمجموعة من الفروقات كل منها يساوي الفرق بين الدرجة الحقيقية والدرجة المتحصل عليها، والخطأ المعياري له علاقة مباشرة بثبات الاختبار، فكلما كانت درجة ثبات الاختبار عالية كلما قل خطأ القياس المعياري. ويمكن تعريف الخطأ المعياري للقياس أيضاً بأنه الانحراف المعياري لدرجات طالب ناتجة عن تكرار إجابته للاختبار نفسه (أو اختبار مماثل له)، ولأن مثل هذه البيانات لا تتوافر عادة فإنها تؤخذ من اختبار مجموعة واحدة من الطلبة. يمكن حساب الخطأ المعياري للاختبار بمعرفة تباين (Variance) الدرجات وثبات الاختبار، فإذا حصل طالب على درجة مقدارها (90) في اختبار خطأه المعياري يساوي (3) فهذا يعني أن هناك احتمالاً قدره (68%) أن تكون الدرجة الحقيقة للطالب واقعة في فترة الثقة بين (87 و93).

- درجة الصف المعيارية (Grade Norm): الدرجة المتوسطة التي حصل عليها طالب مصنف في مستوى دراسي محدد.

- درجة معيارية (Normed Score): هي الدرجة التي تمكن من مقارنة أداء طالب بأداء عينة عامة ممثلة من الطلبة، يتم غالباً تحويل الدرجات الخام التي يحصل عليها الطلبة إلى درجات معيارية لغرض ترجمة وفهم أداء الطلبة.

- قيم معيارية (Normative values): توزيع إحصائي لقيم مقننة يحدد المرجعية التي يتقرر في ضوئها معنى درجة الاختبار التي يحصل عليها طالب أو مجموعة محددة من الطلبة، واحتساب القيم المعيارية يتم في ضوء الأداء الحقيقي لطلبة المجموعة القياسية من مختلف الأعمار والفصول الدراسية،

ونظراً لأنها أي القيم المعيارية (Norms) تعبر عن الأداء المتوسط فيجب ألا ينظر إليها باعتبارها الأداء المستهدف المرغوب فيه، وأكثر القيم المعيارية شيوعاً هي: الانحراف المعياري، والرتبة المئينية.

- مجموعة معيارية (Norm Group): عندما يتم بناء وتطوير الاختبارات فإنها تعطي في البداية لمجموعة كبيرة من الطلبة (يقدر حجمها بالآلاف في حالة الاختبارات التحصيلية) وتحتسب عندئذ كل المؤشرات (البارمترات) الإحصائية الممكنة لتستخدم في بناء معايير التصحيح، وتشكل هذه المجموعة الطلابية الكبيرة ما يسمى بالمجموعة المعيارية، وعندما يعطى الاختبار في وقت لاحق لمجموعة محدودة من الطلبة فإن أداءهم على هذا الاختبار يتم تفسيره ومقارنته في ضوء نتائج طلبة المجموعة المعيارية، وفي ضوء نتائج طلبة هذه المجموعة المحدودة، وكذلك نتائج المجموعة المعيارية يتم تصنيف أداء كل طالب على أنه متوسط أو فوق المتوسط أو دون المتوسط، والمجموعة المعيارية هي عينة من الطلبة تم اختبارها لغرض احتساب الأداء القياسي (Norms) على اختبار مقنن.

- معايرة (Calibration): هي عملية تستخدم لبناء مقياس (Scale) لتمثيل درجات الاختبار بما في ذلك المتوسط الحسابي والانحراف المعياري وشكل توزيع الدرجات، بحيث يمكن مقارنة درجات الطلبة على أي مقياس آخر مماثل.

وبناءً على ما سبق فإن معايير الاختبار خطوة مهمة من خطوات إعداد الاختبار وتصميمه، كما يُعد شرطاً أساسياً من شروط الاختبار الجيد يعتمد على تحويل الدرجات الخام إلى درجات معيارية أو العمر التحصيلي أو النسبة التحصيلية، والتي يمكن ترجمتها إلى مستويات محددة للتحصيل، فمثلاً يصعب تفسير الدرجات الخام في غياب إطار مرجعي يتم على أساسه تفسير هذه الدرجات، فالدرجة (115) التي قد يحصل عليها المفحوص في اختبار للذكاء لا تعني شيئاً ما لم تنسب الى معيار أو محك معين، لأنها قد تعكس صورة خاطئة عن ترتيب المفحوص بالنسبة لأقرانه. إن هذا يدل على عدم إمكانية تفسير درجات الاختبارات والمقاييس المقننة تفسيراً مطلقاً كما هو الحال في القياس الفيزيقي، وإنما يكون بإرجاع الدرجة إلى معيار الجماعة أي أنه تفسير نسبي.

ثانياً: الدراسات السابقة: توصلت العديد من الدراسات إلى أن هناك قصوراً واضحاً في الاختبارات التي يعدها المعلمون وفق المعايير المحددة لهذا الغرض، فقد أجريت العديد من الدراسات التي اهتمت بتحليل أسئلة الاختبارات التحصيلية لبعض المواد الدراسية بهدف تشخيص وضع هذه الاختبارات، لمعرفة مدى مطابقتها للمعايير اللازمة لإعداد الاختبارات الجيد من أجل تطوير ورقة الاختبار من حيث الشكل والمضمون، ومراعاة الأسس والمعايير التي يجب أن تكون عليها هذه الورقة، واهتمت بعض الدراسات بمعرفة مدى امتلاك المعلمين لقواعد بناء الاختبارات التحصيلية وممارساتهم، كما اهتم بعضها بتقييم صلاحية الاختبارات وفعاليتها، ومن بين هذه الدراسات:

استهدفت دراسة القحطاني (1996) التعرف إلى طبيعة الأسئلة التي يقوم بإعدادها معلمو الدراسات الاجتماعية، وكذلك الأسئلة الواردة في كتب الدراسات الاجتماعية في منطقة أبها التعليمية لمعرفة ماهية مستوى هذه الأسئلة وتحديد نواحي القوة والضعف فيها. وأجريت هذه الدراسة على المرحلة المتوسطة، حيث قام الباحث بجمع الاختبارات النهائية للفصلين الأول والثاني، من إحدى عشرة مدرسة وقد أظهرت نتائج الدراسة أن الاختبارات المعدة من قبل المعلمين، كانت حسب تصنيف بلوم في مستوى التذكر ثم

تتدرج نسبة الأسئلة في الانخفاض في مستوى الاستيعاب حتى مستوى التحليل أما أسئلة التركيب والتقويم فقد كانت غير واردة.

كما استهدفت دراسة الهياجنة (1998) تقويم الأسئلة الكتابية الاختبارية التي يعدها معلمو اللغة العربية في المرحلة الثانوية وفق تصنيف بلوم إلى الكشف عن مستويات الأسئلة الكتابية الاختبارية التي يعدها معلمو اللغة العربية في مرحلة التعليم الثانوي في محافظة إربد وفق تصنيف بلوم تكونت عينة الدراسة من (104) من المعلمين والمعلمات، منهم (46) معلماً، و(58) معلمة، وتم استخدام دليل مستويات أهداف المجال المعرفي وفق تصنيف بلوم، وكانت نتائج الدراسة: استخدام معلمو عينة الدراسة أنماطاً مختلفة من الأسئلة الكتابية الاختبارية ضمن المستوى الواحد من مستويات بلوم لأهداف المجال المعرفي. وعدم وجود فروق في ذلك تبعاً لمتغيرات: الجنس والتخصص، والمؤهل والخبرة.

وأجرى الجنازرة (1999) دراسة هدفت إلى تقويم الاختبارات التحصيلية التي يعدها المعلمون والمعلمات في مبحث الكيمياء خراج الاختبار الجيد، في القدس وتكون مجتمع للصف العاشر وفق معايير تصميم والدراسة من الاختبارات التحصيلية لجميع معلمي الكيمياء في المدارس الحكومية والبالغ عددهم (97) معلماً ومعلمة، بينت النتائج أن مستوى تقويم المعلمين للاختبارات كان بمستوى مرتفع، وخلصت النتائج بوجود فروق بين المعلمين والمعلمات في تقويم الاختبارات لصالح الذكور، بينما عدم وجود فروق في ذلك وفق متغيرات والخبرة والمؤهل التربوي.

أجرى شبانه (2000) دراسة هدفت إلى تقييم الاختبارات التحصيلية في مادة الرياضيات لطلبة الصف العاشر في منطقة لواء واد السير في الأردن، حيث قام الباحث بإعداد امتحان خاص أسماه "الرائز" وطبقه على طلبة عينة الدراسة وهم طلبة الصف العاشر. وكان من أهم نتائج الدراسة وجود تفاوت في مستوى صعوبة أسئلة الامتحانات التي يقوم المعلمون بإعدادها، فبعض الاختبارات كانت سهلة وبعضها كان ذو صعوبة لا تراعي الفروق الفردية بين الطلبة مما يؤثر على نتائج الطلبة.

وهدفت دراسة الكحلوت (2004) إلى تقويم اختبارات اللغة العربية للصف الثالث الأساسي، من خلال معرفة مدى توافر معايير الاختبار الجيد في هذه الاختبارات، حيث أجرت الباحثة تقويماً شاملاًَ للاختبارات المدرسية تناولت شكل الاختبار، ومضمونه، بالإضافة إلى تقويم كل نوع من أنواع الأسئلة الموضوعية والمقالية، وتوصلت نتائج الدراسة إلى أن هناك اهتماماً كبيراً من قبل المعلمين بمعايير شكل الاختبار أكثر من اهتمامهم بمعايير المضمون، وخلو جميع الاختبارات من مفتاح التصحيح، وعدم توافر عدد لا بأس به من المعايير الخاصة بأنماط الأسئلة، إضافة إلى خلو الاختبارات من أسئلة الصواب والخطأ وأسئلة المقابلة.

وهدفت دراسة علوان (2006) إلى تحليل الأسئلة المتضمنة في كتابي المطالعة والأدب والنقد للصف الحادي عشر في المنهاج الفلسطيني الجديد حسب تصنيف بلوم للأهداف المعرفية، وتحديد أنماط التفكير التي تطلبها هذه الأسئلة (تقاربي- تباعدي)، وقد أظهرت نتائج الدراسة أن أسئلة الكتاب تتناول المستويات المعرفية بشكل غير متواز، حيث ركزت على بعض المستويات مثل الفهم والتذكر والتحليل بينما قل التركيز بالنسبة لبعض المستويات(التطبيق، والتركيب، والتقويم)، أما بالنسبة لأنماط التفكير فكان التفكير التقاربي هو الأكثر انتشاراً، إذ وصلت نسبته (72.5 %)، بينما نسبة أسئلة التفكير التباعدي كانت حوالي (27.5%)، أما تنوع الأسئلة بين المقالية والموضوعية فقد أشارت النتائج إلى أن النوع السائد هو الأسئلة المقالية.

وهدفت دراسة سنان (2006) إلى معرفة درجة ممارسة معلمات الرياضيات في المرحلة المتوسطة لمهارات بناء الاختبارات التحصيلية، واختلافها باختلاف المؤهل العلمي وسنوات الخبرة والدورات التدريبية في مجال القياس، ولتحقيق أهداف الدراسة تم استخدام الاختبار الذي أعده جرادات (1988) المكون من (90) مفردة من نوع الاختيار من تعدد لكل مفردة (4) بدائل واستبانة من إعداد جرادات (1988)لقياس درجة الممارسة مكونة من (25) مفردة، وتكونت عينة الدراسة من (93) معلمة. وبينت نتائج الدراسة أن هناك تدنى ملحوظ في معرفة وممارسة معلمات الرياضيات بالمرحلة المتوسطة بمدينة مكة المكرمة لقواعد بناء الاختبارات التحصيلية. كما بينت النتائج وجود فروق جوهرية في مستوى ممارسة المعلمين لمهارات بناء الاختبارات وفق متغير الجنس لصالح الإناث، وعدم وجود فروق في ذلك وفق متغيرات: المؤهل العلمي وسنوات الخبرة والدورات التدريبية.

أما دراسة سيف (2008) فهدفت التعرف إلى مستوى معرفة المشرفين التربويين بمبادئ بناء الاختبار بمحافظة تعز، وأثر كل من الخبرة والتخصص ومكان العمل فى مستوى معرفة المشرفين بمبادئ بناء الاختبارات التحصيلية، على عينة مكونة من (296) مشرفاً ومشرفة، وبينت نتائج الدراسة أن هناك تدنى فى مع رفة المشرفين التربويين بمبادئ بناء الاختبارات التحصيلية، كما بينت النتائج عدم وجود فروق دالة إحصائياً فى مستوى معرفة المشرفين بمبادئ الاختبارات التحصيلية تعزى للخبرة أو التخصص أو مكان العمل.

وأجرى مولتون (Moulton, 2013) دراسة هدفت إلى التعرف على مدى توزيع أسئلة عينة من الاختبارات على الأهداف المعرفية، ومدى تقييم المعلمين للاختبارات وفق المعايير اللازمة. تكونت العينة من (78) معلماً ومعلمة، وأظهرت النتائج ارتفاع تقيم المعلمين للاختبارات وفق المعايير اللازمة لبناء الاختبارات، كما توصلت الدراسة إلى أن الأسئلة تركزت على الأهداف الدنيا، كما توصلت الدراسة إلى عدم وجود فروق جوهرية فى نسبة توزيع هذه الأسئلة تعزى لمتغيرات الجنس، والمرحلة، والتخصص، والخبرة.

وأجرى فلمينج وشامبرز (Fleming & Chambers, 2015) دراسة هدفت إلى تحليل أسئلة الاختبارات التي يعدها المعلمون داخل الصف، وقد استخدم الباحثان تصنيف بلوم للمجال المعرفي معياراً للتحلي، وباستخدام النسب المئوية لتحليل البيانات تبين أن (80٪) من الأسئلة المشمولة بالدراسة تقع ضمن مستوى التذكر حيث تقيس مقدرة الطالب على تذكر الحقائق والمصطلحات والقواعد والمبادئ، أما الأسئلة التي تقيس القدرة على التطبيق والتحليل فقد كانت بنسب قليلة، أما المستويات الأُخرى فلم تشمل أي نسبة تذكر. ومن جهة أخرى أظهرت النتائج أن مستوى تقييم الاختبارات كان مرتفعاً، ووجود فروق في هذا المستوى تبعاً لمتغير الجنس والمرحلة لصالح الإناث ومعلمو المرحلة الثانوية، بينما عدم وجود فروق في ذلك وفق متغيرات: التخصص والخبرة والمؤهل.

أما دراسة الفرجات (2015) والتي هدفت التعرف إلى مدى كفاءة معلمي المدارس الحكومية الثانوية التابعة لمديرية التربية والتعليم بمحافظة العقبة في بناء الاختبارات التحصيلية وفق معايير الاختبار الجيد، والوقوف على مستوى الكفاءة المعرفية لديهم في مجال الاختبارات والمقاييس، ودرجة ممارستهم للكفايات المتعلقة ببناء الاختبارات التحصيلية وفق معايير الاختبار الجيد، حيث طبقت الدراسة على عينة مكونة من (315) معلماً ومعلمة، ومن أهم النتائج التي توصلت إليها الدراسة أن مستوى معرفة المعلمين في مجال بناء الاختبارات كانت متوسطة، ومستوى ممارسة المعلمين للكفايات المتعلقة ببناء الاختبارات التحصيلية

كانت متدنية. ووجود فروق دالة في مستوى بناء الاختبارات وفق معايير الاختبار الجيد وفق متغيري الجنس لصالح المعلمات، والمرحلة لصالح الثانوية، وعدم وجود فروق في ذلك وفق متغيرات: المؤهل والتخصص والخبرة والعمر.

وقام بفيفر ودافس (Pfeiffer & Davis, 2016) هدفت إلى تحليل أسئلة اختبارات مواد التربية الاجتماعية والأحياء والعلوم والاقتصاد المنزلي للصف التاسع الأساس وفق معايير بناء الاختبار الجيد. وكان من أهم نتائج هذه الدراسة احتلت أسئلة التذكر في مواد التربية الاجتماعية ومادة الأحياء المرتبة الأولى، وأن مستوى تقييم المعلمين كان متوسطاً، وعدم وجود فروق في ذلك وفق متغيرات: الجنس والمرحلة والتخصص والمؤهل والخبرة.

وأجرى ريزنر (Risner, 2017) دراسة بهدف إلى تحليل المستويات المعرفية حسب تصنيف بلوم في كتب العلوم للصف الخامس الأساسي، وذلك للتعرف إلى مدى وجود فروق بين توزيعه الأسئلة على مستويات بلوم المعرفية الستة وقد أشارت النتائج إلى أن(61%) من الأسئلة هي في المستوى المعرفي (34%) استيعاب (5%) تطبيق وسؤال واحد فقط كان في مستوى التقييم ولم يوجد هناك أي أسئلة في مستوى التحليل والتركيب.

وقام هنتر (Hinter, 2017) بدراسة هدفت إلى قياس مستويات التفكير العليا عند طلبة الصفين الرابع والسادس الأساسيين، واستخدم الباحث سبعة اختبارات للمواد الدراسية التي يدرسها للطلبة لتحقيق هذا الهدف، واعتمد الباحث على تصنيف بلوم لمستويات المجال المعرفي لتحليل الأسئلة. وأظهرت نتائج الدراسة أن ستة اختبارات من عينة الدراسة لا تقيس مهارات التفكير العليا عند الطلبة في حين كان اختبار الرياضيات الأكثر تحقيقاً لذلك. كما أظهرت النتائج أن مستوى تقييم المعلمين لمستوى الاختبار كان منخفضاً، وعدم وجود فروق جوهرية في مستوى هذا التقييم وفق متغيرات: الجنس والمؤهل والخبرة.

وهدفت الدراسة دعائكة وجخراب وجعفور (2018) التعرف إلى مستوى معرفة معلمي التعليم الثانوي لكفايات إعداد الاختبارات التحصيلية الجيدة في ضوء أدوارهم الجديدة، على عينة مقدرة بـ(232) معلماً وعلمة من ثانويات ولاية الوادي وسط، وتوصلت الدراسة إلى عدد من النتائج من أهمها: أن مستوى معرفة أساتذة التعليم الثانوي لكفايات إعداد الاختبارات التحصيلية الجيدة في ضوء أدوارهم الجديدة كانت منخفضة. وعدم وجود فروق جوهرية في هذا المستوى للتقييم وفق متغيرات: الجنس والتخصص والمؤهل والخبرة.

مصطلحات الدراسة:

تتمحور الدراسة الحالية حول المفاهيم والمصطلحات الإجرائية الآتية:

- **المعايير**: هي ما يتم القياس بناءً عليها وهي الإنجاز المخطط أو المستهدف، ويتم التعبير عنه بشكل كمي, وهو يوضح إلى أي مدى يتم السير في بناء الاختبار الجيد من أجل تحقيق الأهداف التعليمية بشكل صحيح. وتعرف المعايير إجرائياً في الدراسة الحالية بالدرجة التي يحصل عليها المفحوص على الأداة المعدة لهذا الغرض.

- **مفهوم الاختبار**: كلمة اختبار عادة ما توحي في الذهن أنه عبارة على سلسلة من الأسئلة المقننة التي تعرض على شخص معين ويطلب منه الإجابة عنها كتابة أو شفهياً، إلا أن هناك بعض الاختبارات التي لا تتطلب من المفحوص إجابة معينة وإنما تتطلب منه أداءً حركياً أو مجموعة من الأداءات الحركية على آلة معينة كاختبار قيادة السيارة في الشارع؛ فالاختبارات من أهم الأدوات المستخدمة لتقويم الدارسين

وتكاد تطغى على غيرها من الوسائل التربوية الأخرى، ويعرف كرونباخ (Cronbach, 1997, 18) الاختبار بأنه "طريقة منظمة لمقارنة سلوك شخصية بسلوك شخصية أخرى"، ويعرفه أنكوف (Angoff, 1982, 39) بأنه "الأداة التي تستخدم في قياس المعرفة والفهم والمهارة في مادة دراسية وتدريبية معينة أو مجموعة المواد"، مما سبق يتضح أن الاختبار عبارة عن عينة مختارة يقررها واضع الاختبار وفق مجموعة من الأسس والقواعد مبنية على المادة الدراسية في أساسه وفي خطوطها العريضة وتتضمن مجموعة من الأسئلة أو المثيرات التي يتعرض لها الدارس، وإجرائياً يعرف الاختبار الجيد في الدراسة الحالية بالتزام المعلمين بمعايير الاختبار الجيد والتي تشتمل عليها الأداة المعدة لهذا الغرض.

حدود الدراسة:

تعمم نتائج الدراسة في نطاق الحدود الآتية:

الحد الزماني: أجريت هذه الدراسة في الفصل الدراسي الأول من العام الدراسي (2020\2019).

الحد المكاني: أجريت هذه الدراسة في المدارس الحكومية الأساسية والثانوية في المدارس الحكومية بمحافظة طولكرم في فلسطين.

الحد البشري: تمثلت عينة الدراسة من معلمي مرحلتي التعليم الأساسية والثانوية في محافظة طولكرم.

الحد الموضوعي: تتحدد نتائج الدراسة ويمكن تعميمها في ضوء متغيراتها والأداة المعدة لقياس معايير بناء الاختبار الجيد وفي نطاق خصائصها السيكومترية (الصدق والثبات) التي أمكن التحقق منها.

الطريقة والإجراءات

أولاً: منهج الدراسة: في ضوء طبيعة الدراسة والبيانات المراد الحصول عليها استخدم الباحث المنهج الوصفي التحليلي، فهو المنهج المناسب لهذه الدراسة، ويحقق أهدافها بالشكل الذي يضمن الدقة والموضوعية.

ثانياً: مجتمع الدراسة: تكون مجتمع الدراسة من معلمي المرحلتين الأساسية والثانوية في المدارس الحكومية بمحافظة طولكرم والبالغ عددهم (1891) معلماً ومعلمة، تبعاً لإحصائيات وزارة التربية والتعليم الفلسطينية للعام الدراسي (2020-2019)، والموزعين حسب الجدول (1) الآتي:

جدول (1): توزع مجتمع الدراسة وفق متغيري الجنس والمرحلة

الجنس / المرحلة	الذكور		الإناث		المجموع	النسبة الكلية
	العدد	النسبة	العدد	النسبة		
المرحلة الأساسية	769	.7 %40	843	.6 %44	1612	.1 85 %
المرحلة الثانوية	129	.8 %6	150	.9 %7	279	.7 %14
المجموع	**89**	**.5**	**993**	**.5**	**1**	**%100**

	891	**%52**		**%47**	**8**	

ثالثاً: عينة الدراسة:

تكونت عينة الدراسة من (170) معلماً ومعلمة من المرحلتين الأساسية والثانوية في المدارس الحكومية بمحافظة طولكرم، وهم يمثلون ما نسبته (9%) من المجتمع الأصلي، تم اختيارهم بطريقة عشوائية طبقية تبعاً لمتغيري الجنس والمرحلة التعليمية، وهم موزعون تبعاً لمتغيرات الدراسة كما هو مبين في الجدول (2) الآتي:

الجدول (2): توزيع عينة الدراسة تبعاً لمتغيراتها المستقلة

المتغيرات المستقلة	مستويات المتغير	العدد	النسبة المئوية
الجنس	ذكر	81	47.6
	أنثى	89	52.4
المرحلة التعليمية	أساسية	145	85.3
	ثانوية	25	14.7
سنوات الخبرة	(5) سنوات فأقل	27	15.9
	من (6 -10) سنة	33	19.4
	من (11-15) سنة	45	26.5
	(16) سنة فأكثر	65	38.2
المؤهل العلمي	دبلوم	30	17.6
	جامعي	120	70.6
	دراسات عليا	20	11.8
نوع التخصص	مواد علمية	48	28.2
	مواد أدبية	108	63.6
	مواد تجارية	14	8.2

رابعاً: أداة الدراسة: بعد إطلاع الباحث على عدد من الدراسات السابقة والأدوات المستخدمة فيها (Black, 2018؛ Fleming & Chambers, 2018؛ ساعد وعامر، 2017؛ Hinter, 2017؛ Risner, 2017؛ Friedlend & Folt, 2015؛ Frances, 2014؛ Serka, 2013؛ حماد، 2011؛ الخوالدة والمشاعلة والقضاة، 2007؛ شبانة، 2000؛ الجنازرة، 1999؛ الهياجنة، 1998)، قام ببناء استبانه خاصة من أجل التعرف إلى أهمية معايير الاختبار الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية في المدارس الحكومية في محافظة طولكرم، وقد تكونت الأداة في صورتها النهائية من جزأين: الأول تضمن بيانات أولية عن المبحوثين تمثلت في الجنس، والمرحلة التعليمية، وسنوات الخبرة، والمؤهل العلمي، ونوع التخصص، أما الثاني فقد تكون من الفقرات التي تقيس مدى تأثير المخاوف

السائدة لدى طلبة المرحلة الأساسية وأساليب علاجها والوقاية منها من وجهة نظر المعلمين في مدينة طولكرم، حيث بلغ عدد هذه الفقرات (38) فقرة، كل منها يعبر عن معيار مهم من معايير بناء الاختبار الجيد. كما تم تصميم الاستبانة على أساس مقياس ليكرت خماسي الأبعاد وقد بنيت الفقرات بالاتجاه الإيجابي والسلبي، وأعطيت الأوزان للفقرات كما يلي: موافق جداً (5) درجات، وموافق (4) درجات، ومحايد (3) درجات، ومعارض (2) درجتين، ومعارض جداً (درجة واحدة). وبذلك فإن الدرجة المرتفعة على الاستبانة تشير إلى ارتفاع مستوى الالتزام بمعايير الاختبار الجيد لدى المعلمين، بينما تشير الدرجة المنخفضة عليها إلى انخفاض مستوى هذا الالتزام،

صدق المقياس وثباته: للتأكد من صدق الاستبانة استخدمت طريقتين هما: الأولى؛ صدق المحكمين (Arbitration Validity) من خلال عرض الاستبانة بصورتها الأولية على عدد من المحكمين ذوي الخبرة والاختصاص في مجالات تربوية ونفسية مختلفة بلغ عددهم (9) محكمين، طلب منهم إبداء الرأي حول فقرات المقياس بهدف التأكد من مناسبتها للمفهوم المراد قياسه، وقد تم الأخذ بملاحظاتهم عند اخراج المقياس بصورته النهائية، حيث تم حذف فقرتين بسبب تكرار موضوعهما، وبذلك أصبحت الاستبانة بصورتها النهائية مكون من (36) فقرة. والطريقة الثانية؛ هي صدق البناء (Construct Validity) باستخدام معادلة بيرسون إذ تم حساب معاملات الارتباط بين درجات أفراد العينة الاستطلاعية للدراسة وحجمها (55) معلماً ومعلمة لم يدخلوا في عينة الدراسة الفعلية على الفقرات والمجموع الكلي للمقياس، وقد تراوحت هذه المعاملات ما بين (30. 0 – 79. 0)، وجميع هذه المعاملات دالة إحصائياً كما هو مبين في الجدول (3) الآتي:

الجدول (3): قيم معاملات ارتباط الفقرات والمقياس ككل

الرقم الفقرة	معامل الارتباط	الرقم الفقرة	معامل الارتباط	الرقم الفقرة	معامل الارتباط	الرقم الفقرة	معامل الارتباط	الرقم الفقرة	معامل الارتباط
1	.56**0	8	.37*0	15	.51**0	22	.34*0	29	.69**0
2	.64**0	9	.63**0	16	.41**0	23	.47**0	30	.43**0
3	.31*0	10	.43**0	17	.65**0	24	.69**0	31	.67**0
4	.61**0	11	.48**0	18	.69**0	25	.79**0	32	.39*0
5	.51**0	12	.44**0	19	.51**0	26	.53**0	33	.34*0
6	.68**0	13	.77**0	20	.41**0	27	.67**0	34	.65**0

7	.31 *0	1 4	.30 *0	2 1	.65 **0	2 8	.44 *0	3 5	.69 **0
								3 6	.44 **0

- دالة عند مستوى الدلالة ($\alpha \leq 0.05$)
- * دالة عند مستوى الدلالة ($\alpha \leq 0.01$)

كما تم التحقق من ثبات المقياس باعتماد طريقة الاتساق الداخلي (Internal Consistency) وذلك باستخدام معادلة كرونباخ ألفا (Cronbach -Alpha)، باستخدام العينة الاستطلاعية السابقة، وقد بلغت قيمة معامل الثبات العام للمقياس (0. 88)، وقد اعتبر الباحث ذلك مؤشرات مناسبة لصدق الأداة وثباتها. ولتفسير استجابات المفحوصين على هذه الاستبانة اعتمد المعيار الوسطي الآتي:

- (2. 33) فأقل — مستوى منخفض للمعايير
- (3 .66 – 2 .34) — مستوى متوسط للمعايير
- أكثر من (3. 66) — مستوى مرتفع للمعايير

المعالجات الإحصائية:

من أجل معالجة البيانات تم استخدام البرنامج الإحصائي للعلوم الاجتماعية (spss) وذلك باستخدام المعالجات الإحصائية الآتية:

1. المتوسطات الحسابية والانحراف المعياري والنسب المئوية.
2. اختبار (ت) لمجموعتين مستقلتين (Independent-Samples t-Test).
1. 3. تحليل التباين متعدد المتغيرات (3×3×4) (Multivariat Analysis of Variance -3).
2. معادلة كرونباخ - ألفا (Cronbach – Alpha Formula).

النتائج ومناقشتها

أولا : النتائج المتعلقة بالسؤال الأول وهو: ما معايير الامتحان الجيد من وجهة نظر معلمي المرحلة الأساسية والثانوية في المدارس الحكومية في محافظة طولكرم؟

للإجابة عن السؤال حسبت المتوسطات الحسابية والانحراف المعياري والنسب المئوية لكل فقرة من فقرات الاستبانة والمتوسط الكلي والتقييم لأهميتها كما هو مبين في الجدول (4) الآتي:

جدول(4): المتوسطات الحسابية والانحراف المعياري والنسب المئوية وتقييم أهمية معايير الاختبار الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية

الرتبة	الرقم التسلسلي	الفقرات	المتوسط الحسابي	الانحراف المعياري	النسبة المئوية	التقييم
1	15	درجة وضوح الأسئلة وسهولة لغة الاختبار	4.57	0.57	91.4	مرتفع
2	2	طريقة توزيع الدرجات على	4.55	0.57	91.0	مرت

		أسئلة الاختبار مناسبة				فع
3	27	التنوع في صعوبة الأسئلة ومراعاة الفروق الفردية	4.53	0.73	90.6	مرتفع
4	14	أسئلة الاختبار متوافقة مع ما تم شرحه	4.52	0.77	90.2	مرتفع
5	1	تطابق الاختبارات مع وصف ومفردات المقرر	4.42	0.69	88.4	مرتفع
6	7	يشتمل الاختبار على أسئلة تميز الطلبة المتفوقين	4.39	0.76	87.8	مرتفع
7	33	يغطي الاختبار المستويات المعرفية المختلفة	4.38	0.86	87.6	مرتفع
8	11	يكتب بصيغة تتناسب مع المرحلة الأكاديمية للمتعلم	4.37	0.71	87.4	مرتفع
9	4	تطابق الاختبار مع الوقت المخصص	4.37	0.74	87.4	مرتفع
10	5	توزيع أسئلة الامتحان على المادة العلمية يحقق شمولية الاختبار	4.26	0.81	85.2	مرتفع
11	13	الدرجة الكلية المعطاة على الاختبار النهائي مناسبة	4.25	0.87	85.0	مرتفع
12	16	الوقت الكلي المخصص للاختبار كاف وغير مجهد	4.20	0.93	84.0	مرتفع
13	8	يحقق العدالة بالنسبة لتوزيع الدرجات على الطلبة	4.17	0.95	83.4	مرتفع
14	17	موعد الاختبار كان مناسباً من حيث علاقته بالاختبارات الأخرى	4.15	0.93	83.0	مرتفع
15	6	يحتوي الاختبار على أجزاء تقيس مهارات التفكير العليا	4.10	0.75	82.0	مرتفع
16	20	ضبط إجراءات التصحيح من خلال الإجابة النموذجية	4.08	0.84	81.6	مرتفع
17	31	يتوافق مع مستوى تفكير الطلبة	4.07	0.90	81.4	مرتفع
18	10	تظهر علامات الترقيم بوضوح وبشكل سليم على الاختبار	4.05	0.93	81.0	مرتفع

19	13	مراعاة سلامة اللغة في صياغة أسئلة الاختبار	3.99	0.75	79.8	مرتفع
20	26	تقيس الأسئلة نواتج التعلم المستهدفة	3.99	0.84	79.8	مرتفع
21	29	مراعاة شروط صياغة الأسئلة بأنواعها المختلفة	3.99	0.90	79.8	مرتفع
22	23	مراعاة الظروف النفسية للطلبة كقلق الاختبار والتوتر	3.98	0.92	79.6	مرتفع
23	22	تحقق الأهداف التربوية في التربية المدرسية	3.97	0.91	79.4	مرتفع
24	12	تخلو من الاشارات والتلميحات التي توحي بالإجابة الصحيحة	3.97	0.95	79.4	مرتفع
25	18	يرتبط بالحياة العملية الواقعية ولم يركز على الحفظ فقط	3.86	0.84	77.2	مرتفع
26	9	يطبق في قاعة مريحة نفسياً وفيزيقياً	3.86	0.90	77.2	مرتفع
27	24	يوجد تحسين على نوعية التعلم والوصول الى درجة الاتقان	3.86	0.92	77.2	مرتفع
28	19	البعد عن الأسئلة التي تتطلب مجرد استدعاء معلومات	3.85	0.91	77.0	مرتفع
29	21	أسئلة الاختبار متوقعة	3.64	1.01	72.8	متوسط
30	25	تسهم في حل الاختبارات كثير من مشكلات التعليم	3.60	1.08	72.0	متوسط
31	30	تجنب الأسئلة الاختيارية لتتم المقارنة بين جميع الطلبة مقارنة سليمة	3.38	0.84	67.6	متوسط
32	36	وضوح تعليمات الاختبار	3.27	0.98	65.4	متوسط
33	32	تضمن ورقة الاختبار الزمن المتاح للاختبار	3.06	1.09	61.2	متوسط
34	34	وضوح عدد صفحات الاختبار	3.05	0.95	61.0	متوسط
35	35	تحدد درجة كل سؤال من أسئلة الاختبار	3.01	1.16	60.2	متوسط

36	**28**	تتدرج الأسئلة من البسيط إلى الأصعب	3.00	1.14	60.0	متوسط
المتوسط الكلي			**3.91**	**0.42**	**78.2**	**مرتفع**

يوضح الجدول (4)، أن تقديرات أفراد الدراسة كانت بمستوى مرتفع على الفقرات (15 , 2 , 27 , 14 , 1 , 7 , 33 , 11 , 4، 5 ، 13 , 16 , 8، 17 , 6 , 20 , 31، 10 ، 3، 26، 29، 23، 22، 12، 18، 9، 24، 19)؛ حيث تراوحت النسبة المئوية للاستجابة عليها ما بين (77% - 4. 91%), كما كانت هذه التقديرات بمستوى متوسط على باقي الفقرات؛ حيث تراوحت النسبة المئوية للاستجابة عليها ما بين (60%-72.8 %). أما بالنسبة للمتوسط الكلي لمعايير الامتحان الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية فقد كان بمستوى مرتفع، إذ بلغت النسبة المئوية للاستجابة (2. 78%). وعند مقارنة هذه النتيجة مع الدراسات السابقة تبين أنها تتفق مع دراسات (الجنازرة، 1999؛ الكحلوت، 2004؛ Moulton, 2013؛ Fleming & Chambers, 2015)؛ التي أظهرت نتائجها أن مستوى التزام المعلمين بمعايير اعداد الاختبار الجيد كان مرتفعاً. بينما تعارضت مع دراسات (القحطاني، 1996؛ شبانة، 2000؛ سنان، 2006؛ سيف، 2008؛ الفرجات، 2015؛ Pfeiffer & Davis, 2016؛ Hinter, 2017؛ دعائكة وجخراب وجعفور، 2018)؛ التي أظهرت نتائجها أن مستوى التزام المعلمين بمعايير اعداد الاختبار الجيد كان متوسطاً أو منخفضاً.

وقد يعود ذلك إلى إدراك المعلمين إلى أهمية إعداد اختباراتهم وتطويرها بما يناسب المعايير المعمول بها وفق آليات وأساليب القياس والتقويم، أو معرفتهم الجيدة بالمفاهيم والإجراءات اللازمة لبناء الاختبارات سواء كان ذلك ما يخص الجانب الشكلي للاختبار: كمراعاة سلامة اللغة، واستخدام علامات الترقيم، ومناسبة الوقت المتاح للاختبار، ووضوح الفقرات وتنوعها، وإجراءات التصحيح والعدالة والموضوعية في منح الدرجات، وتطبيقها في ظروف ومواعيد مناسبة، وشروط صياغة الفقرات بأشكالها المختلفة الموضوعية والمقالية، ومراعاة الظروف النفسية للطالب وقت الاختبار كأن تكون مريحة نفسياً وتجنب القلق والتوتر. أم جانب المضمون: كمراعاة التنوع في صعوبة الفقرات، ومناسبتها للمحتوى المراد قياسه، والمستويات المعرفية المختلفة للأهداف، وتراعي الطلبة المتميزين، وتناسب المرحلة العمرية للطلبة، وتقيس المهارات العليا في التفكير.

وتشير استجابات عينة الدراسة الحالية إلى كفايات المعلمين إجمالاً بمعايير الاختبار الجيد، فهم من جهة لديهم دراية بأهمية تحديد الغرض من الاختبار قبل العمل على تنفيذه كخطوة أولى لبناء اختبار تحصيلي جيد، كما تعد كفاية تحليل محتوى من بين الكفايات التي كانت استجابة عينة الدراسة عليها مرتفعة، مما يعن أن المعلمين لديهم الفهم الكامل لموضوعات المقرر الدراسي وجزئيات كل موضوع وهو ما مكنهم من الالتزام بالمعايير اللازمة لبناء الاختبارات، ومن جهة أخرى فإن لكفايات تحديد الأهداف، والذي قد يعود إلى وعي عينة الدراسة بأن الوصف الجيد للنواتج المقصودة من التدريس يعكس بدقة المطلوب تحقيقه من الاختبار أو قد يعود إلى أن عينة الدراسة تحدد الأهداف التي يقيسها الاختبار حسب أهداف ما تم تدريسه في غرفة الصف، أو مراعاة التوازن وشمولية الأهداف للمستويات

التي يجب أن تمثلها، بالتالي فإن ما ينتج عن التحديد الجيد للأهداف يتبعه التزام كبير باتباع المعايير عند الإعداد الجيد للاختبار.

ثانياً: النتائج المتعلقة بالسؤال الثاني وهو: هل توجد فروق ذات دلالة إحصائية في مستوى الالتزام بمعايير الاختبار الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية تبعاً إلى متغير الجنس؟

للإجابة عن هذا السؤال حسبت المتوسطات الحسابية والانحرافات المعيارية لاستجابات أفراد الدراسة للالتزام بمعايير الاختبار الجيد تبعاً إلى متغير الجنس، كما استخدم اختبار (ت) لمجموعتين مستقلتين، فكانت النتائج كما هو مبين في الجدول (5):

جدول(5): نتائج اختبار (ت) لدلالة الفروق الإحصائية بين استجابات أفراد الدراسة للالتزام بمعايير الامتحان الجيد تبعاً إلى متغير الجنس

ذكر		أنثى		قيمة (ت) المحسوبة	مستوى الدلالة
المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري		
3.77	0.33	4.00	0.44	2.93	*0.004

* دالة عند المستوى الدلالة ($\alpha \leq 0.01$)

يوضح الجدول (5) السابق، وجود فروق ذات دلالة إحصائية عند المستوى ($\alpha \leq 0.01$) في مستوى التزام أفراد الدراسة بمعايير الاختبار الجيد تبعاً إلى متغير الجنس, وذلك لصالح الإناث، بمعنى أن المعلمات قدرن التزامهن بمعايير الاختبار الجيد بمستوى أكبر مقارنة بالمعلمين الذكور. وعند مقارنة هذه النتيجة مع الدراسات السابقة تبين أنها تتفق مع دراسات (سنان، 2006؛ Fleming & Chambers, 2015؛ الفرجات، 2015)؛ التي أظهرت نتائجها وجود فروق في مستوى التزام المعلمين بمعايير اعداد الاختبار الجيد تبعاً إلى متغير الجنس وذلك لصالح المعلمات. بينما تعارضت مع دراسة الجنازرة (1999)؛ التي أظهرت وجود فروق في هذا المستوى لصالح المعلمين الذكور. كما تعارضت مع دراسات (Moulton, 2013؛ Pfeiffer & Davis, 2016؛ Hinter, 2017؛ دعائكة وجخراب وجعفور، 2018)؛ التي أظهرت نتائجها عدم وجود فروق جوهرية في مستوى التزام المعلمين من كلا الجنسين بمعايير اعداد الاختبار الجيد. لربما يعزى ذلك إلى تشابه كمية المعلومات والمعرفة التي حصل عليها كل من المعلمين والمعلمات لأن مستوى المعايير للامتحان الجيد هي معايير يكتسبها المعلم أو المعلمة من وزارة التربية والتعليم وفق برامج تدريبية واحدة يخضع لها المعلم من خلال خضوعه لدورات تأهيلية موحدة لا تختلف باختلاف الجنس.

ثالثاً: النتائج المتعلقة بالسؤال الثالث وهو: هل توجد فروق ذات دلالة إحصائية في مستوى الالتزام بمعايير الاختبار الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية تبعاً إلى متغير المرحلة التعليمية؟

للإجابة عن هذا السؤال حسبت المتوسطات الحسابية والانحرافات المعيارية لاستجابات أفراد الدراسة لمعايير الاختبار الجيد تبعاً إلى متغير المرحلة التعليمية، كما استخدم اختبار (ت) لمجموعتين مستقلتين، فكانت النتائج كما هو مبين في الجدول (6):

جدول(6): نتائج اختبار (ت) لدلالة الفروق الإحصائية بين استجابات أفراد الدراسة للالتزام بمعايير الامتحان الجيد تبعاً إلى متغير المرحلة التعليمية

الأساسية		الثانوية		قيمة (ت) المحسوبة	مستوى الدلالة
المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري		
3.80	0.39	4.00	0.42	2.55	0.01 *

* دالة عند المستوى الدلالة ($\alpha \leq 0.01$)

يوضح الجدول (6) السابق، وجود فروق ذات دلالة إحصائية عند المستوى ($\alpha \leq 0.01$) في مستوى الالتزام بمعايير الاختبار الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية تبعاً إلى متغير المرحلة التعليمية, وذلك لصالح المرحلة الثانوية، بمعنى أن معلمي المرحلة الثانوية قد قدروا التزامهم بمعايير بناء الاختبار بمستوى أكبر مقارنة بمعلمين المرحلة الأساسية. وعند مقارنة هذه النتيجة مع الدراسات السابقة تبين أنها تتفق مع دراسات (Fleming & Chambers, 2015؛ الفرجات، 2015)؛ التي أظهرت نتائجها وجود فروق في مستوى التزام المعلمين بمعايير اعداد الاختبار الجيد تبعاً إلى متغير المرحلة التعليمية وذلك لصالح معلمي المرحلة الثانوية. بينما تعارضت مع دراسات (Moulton, 2013؛ Pfeiffer & Davis, 2016)؛ التي أظهرت نتائجها عدم وجود فروق جوهرية في مستوى التزام المعلمين بمعايير اعداد الاختبار الجيد وفق متغير المرحلة التعليمية. لربما يعزى ذلك إلى أن المرحلة الثانوية يلتزم بها المعلم بالمعايير بمستوى الجيد والذي يتفق مع قرارات الوزارة حول هذه المعايير لهذه المرحلة، لأنه بعد ذلك سوف يتجه الطالب إلى الدراسات الجامعية، وبذلك يتطلب من كل معلم في تلك المرحلة المراجعة والتدقيق حول الاختبارات لكي يحصل الطالب على درجته الحقيقية التي تؤهله للالتحاق بالتخصص المناسب في الجامعة.

رابعاً: النتائج المتعلقة بالسؤال الرابع وهو: هل توجد فروق ذات دلالة إحصائية في مستوى الالتزام بمعايير الامتحان الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية تبعاً إلى المتغيرات: التخصص العلمي والمؤهل العلمي وسنوات الخبرة؟

للإجابة عن هذا السؤال حسبت المتوسطات الحسابية والانحرافات المعيارية لاستجابات أفراد الدراسة للالتزام بمعايير الامتحان الجيد تبعاً إلى المتغيرات: التخصص العلمي والمؤهل العلمي وسنوات الخبرة، فكانت النتائج كما هو مبين في الجدول (7):

الجدول (7): المتوسطات الحسابية والانحرافات المعيارية لاستجابات أفراد الدراسة للالتزام بمعايير الامتحان الجيد تبعاً إلى المتغيرات: التخصص العلمي والمؤهل العلمي وسنوات الخبرة

المتغيرات	المستوى	المتوسط الحسابي	الانحراف المعياري
الجنس	الذكور	3.80	0.69
	الإناث	4.00	0.44
التخصص العلمي	مواد ادبية	3.88	0.74
	مواد علمية	4.04	0.64
	مواد تجارية	3.58	0.60
المؤهل العلمي	دبلوم	3.86	0.66
	بكالوريوس	3.89	0.67
	دراسات عليا	4.07	0.65
سنوات الخبرة	(5) سنوات فأقل	3.98	0.54
	من (6 -10) سنة	3.87	0.80
	من (11-15) سنة	3.88	0.78
	(16) سنة فأكثر	3.97	0.71
المتوسط الكلي		3.91	0.34

يظهر الجدول (7) السابق، وجود فروق ظاهرية بين المتوسطات الحسابية لتقديرات أفراد الدرسة لمستوى الالتزام بمعايير الاختبار الجيد تبعاً إلى متغيرات: التخصص العلمي، والمؤهل العلمي، وسنوات الخبرة للمعلمين، وللتحقق من دلالة الفروق بين هذه المتوسطات استخدم اختبار تحليل التباين متعدد المتغيرات (2×3×3×4)، والمبينة نتائجه في الجدول (8) الآتي:

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة (ف) المحسوبة	مستوى الدلالة
التخصص العلمي	0.017	2	0.009	0.024	0.936
المؤهل العلمي	0.504	2	0.252	0.668	0.186
سنوات الخبرة	1 .018	3	.3390	.8990	0.68
الخطأ (البواقي)	61.111	162	0.377		
المجموع الكلي	1007.744	169			

الجدول (8): نتائج تحليل التباين متعدد المتغيرات(2×3×3×4) لاستجابات التزام أفراد الدراسة بمعايير الاختبار الجيد تبعاً إلى متغيرات: التخصص العلمي، والمؤهل العلمي، وسنوات الخبرة

دال عند مستوى الدلالة ($\alpha \leq 0.01$)

يوضح الجدول (8) السابق، النتائج الآتية:

- عدم وجود فروق ذات دلالة إحصائية في مستوى التزام أفراد الدراسة من المعلمين معايير الامتحان الجيد تبعاً إلى متغير التخصص العلمي, بمعنى أن أفراد الدراسة قد قدروا الالتزام بمعايير الاختبار الجيد بنفس المستوى بغض النظر عن تخصصاتهم علمية كانت أم أدبية أم تجارية. وعند مقارنة هذه النتيجة مع الدراسات السابقة تبين أنها اتفقت مع دراسات (الهياجنة، 1998؛ سنان، 2006؛ سيف، 2008؛ Moulton, 2013؛ Felming & Chambers, 2015؛ الفرجات، 2015؛ Pfeiffer & Davis, 2016؛ Hinter, 2017؛ دعائكة وجخراب وجعفور، 2018)؛ التي أظهرت نتائجها عدم وجود فروق جوهرية في مستوى التزام المعلمين بمعايير اعداد الاختبار الجيد تبعاً إلى متغير التخصص العلمي لديهم. لربما يعزى ذلك إلى تشابه كمية المعلومات والمعرفة التي حصل عليها المعلمين جميعاً بغض النظر عن تخصصاتهم العلمية، وذلك من خلال التدريب قبل المهنة وأثنائها بالتحاقهم بالدورات التدريبية المتخصصة حول بناء الاختبارات شكلاً ومضموناً.

- عدم وجود فروق ذات دلالة إحصائية في مستوى معايير الامتحان الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية تبعاً إلى متغير المؤهل العلمي, بمعنى أن أفراد الدراسة قد قدروا معايير الاختبار الجيد بنفس المستوى بغض النظر عن المؤهل العلمي الذي يحمله. وتعزى هذه النتيجة إلى أن جميع المعلمين على اختلاف مؤهلاتهم التعليمية تسري عليهم الأنظمة المعمول بها من حيث التزامهم بمعايير الامتحان الجيد. وعند مقارنة هذه النتيجة مع الدراسات السابقة تبين أنها اتفقت مع دراسات (الهياجنة، 1998؛ الجنازرة، 1999؛ سنان، 2006؛ Felming & Chambers, 2015؛ الفرجات، 2015؛ Pfeiffer & Davis, 2016؛ Hinter, 2017؛ دعائكة وجخراب وجعفور، 2018)؛ التي

أظهرت نتائجها عدم وجود فروق جوهرية في مستوى التزام المعلمين بمعايير اعداد الاختبار الجيد تبعاً إلى متغير المؤهل العلمي.

- عدم وجود فروق ذات دلالة إحصائية في مستوى معايير الامتحان الجيد من وجهة نظر معلمي المرحلتين الأساسية والثانوية تبعاً إلى متغير سنوات الخبرة للمعلمين, بمعنى أن أفراد الدراسة قد قدروا معايير الاختبار الجيد بنفس المستوى بغض النظر عن سنوات الخبرة لدى هؤلاء المعلمون. وعند مقارنة هذه النتيجة مع الدراسات السابقة تبين أنها اتفقت مع دراسات (الهياجنة، 1998؛ الجنازرة، 1999؛ سنان، 2006؛ سيف، 2008؛ Moulton, 2013؛ Felming & Chambers, 2015؛ الفرجات، 2015؛ Pfeiffer & Davis, 2016؛ Hinter, 2017؛ دعائكة وجخراب وجعفور، 2018)؛ التي أظهرت نتائجها عدم وجود فروق جوهرية في مستوى التزام المعلمين بمعايير اعداد الاختبار الجيد تبعاً إلى متغير سنوات الخبرة لديهم. لربما يعزى ذلك إلى تشابه الأسباب المؤثرة في مستوى الالتزام لكل من المعلمين على اختلاف خبراتهم مثل: مستوى المعرفة الكافية لإعداد الاختبارات قبل الخدمة, واهتمام المشرفين التربويين عند إعطائهم للدورات التدريبية, ومتابعتهم للمعلمين في الميدان. كما يرجع سبب هذه النتيجة أيضاً إلى: أن المعلمين الأقل خبرة يحاولون بتطوير أنفسهم باستمرار باستخدام التقنيات الحديثة، إضافة إلى توفر الرغبة والدافع للعمل في التعليم لحداثة عهدهم بالعمل، وأنهم يعيشون ذات الظروف في العمل والقوانين والتعليمات والأنظمة مع زملائهم ذوي الخبرات الطويلة والاستفادة منهم، ودخولهم في برامج التدريب أثناء الخدمة والاستفادة من الدورات الإشراقية في كيفية اعداد الاختبارات.

خاتمة والتوصيات

تناولت الدراسة الحالية موضوع مدى التزام معلمي المرحلتين الأساسية والثانوية في محافظة طولكم بفلسطين بمعايير بناء الاختبارات الجيد، وقد خلصت إلى مستوى مرتف من الالتزام بهذه المعايير لدى المعلمين بالمرحلتين الأساسية والثانوية، وقد بينت النتائج أن مستوى التزام المعلمات أكبر بهذه المعايير مقارنة بالذكور، وأن معلمي المرحلة الثانوية أكثر التزاماً من معلمي المرحلة الأساسية، وعدم وجود فروق جوهرية في مستوى هذا الالتزام لدى المعلمين وفق متغيرات التخصص والمؤهل العلمي وسنوات الخبرة. وفي ضوء هذه نتائج الدراسة ومناقشتها يمكن التقدم بالتوصيات والمقترحات الآتية:

- ينبغي على المعلم أن يراعي في الاختبار: الاهتمام بورقة الأسئلة إخراجاً وتنظيماً, شكلاً ومضموناً، وصياغة الأسئلة بلغة سليمة لا لبس ولا غموض فيها, أن تكون الأسئلة مناسبة للزمن المخصص لها, وأن تكون الأسئلة متنوعة، وواضحة من حيث التعليمات والزمن المتاح لها، واتباع ترقيم واضح لصفحاتها، ومتدرجة من حيث الصعوبة والسهولة، وتوزيع الدرجات عليها بشكل مناسب.
- على وزارة التربية والتعليم الفلسطينية والمديريات التابعة لها والمتخصصين والمعنيين فيها استخدام برامج التدريب والتأهيل المناسبة لإعداد الاختبارات وتطويرها وفق المعايير المتعارف عليها لبناء الاختبار الجيد.
- على المعلمين الذكور والمعلمين في المرحلة الأساسية بذل المزيد من الجهد للتدرب على بناء الاختبارات واعدادها وفق المعايير الإجرائية من خلال الالتحاق بالدورات التدريبية المتخصصة في القياس والتقويم وتصميم التدريس.
- استخدام الوسائل التقنية في إعداد الاختبارات.

- إجراء مزيد من الدراسات الإجرائية تمس موضوع الدراسة الحالية في علاقته بمتغيرات تربوية ونفسية وشخصية مختلفة.

المراجع العربية

- أبو جراد, حمدي (2011). مدى التزام المعلمين بتحليل نتائج الاختبارات التحصيلية وعلاقته باتجاهاتهم نحوها. مجلة جامعة الأزهر بغزة, سلسلة العلوم الإنسانية, 13(2)، 106-89.
- بديه, أحمد (2009). الامتحانات المدرسية: المشكلة والحل. الجمهورية اليمنية, وزارة التربية والتعليم, عدن: مركز البحوث والتطوير التربوي.
- بركات، زياد (2018). القياس والتقويم النفسي والتربوي بين النظرية والتطبيق. رام الله: عمادة البحث العلمي، جامعة القدس المفتوحة.
- بركات، زياد وصباح، عبد الهادي (2007). مدى تحقيق أسئلة الامتحانات النهائية في جامعة القدس المفتوحة للأهداف التعليمية تبعاً لهرم بلوم. مجلة جامعة القدس المفتوحة للأبحاث والدراسات، (9)، 145-123.
- جعفور، ربيعة (2014). أثر كل من نوع المحكم وطول الاختبار على تحديد درجة القطع لاختبار محكي المرجع يقيس الكفايات الرياضية في العمليات الحسابية على الأعداد بالصف السادس الابتدائي بمدينة جدة. رسالة ماجستير غير منشورة, كلية التربية, جامعة أم القرى, المملكة العربية السعودية.
- الجنازرة، أحمد (1999). تقويم الاختبارات التحصيلية لمادة الكيمياء للصف العاشر وفق معايير إخراج الاختبار التحصيلي الجيد. رسالة ماجستير غير منشورة، جامعة القدس، فلسطين.
- حماد, علي (2011). جودة أسئلة الامتحانات النصفية لبرنامج التربية في جامعة القدس المفتوحة في ضوء معايير جودة الامتحان. ورقة علمية مقدمة إلى اليوم الدراسي "جودة الامتحانات الجامعية" الواقع والمأمول والمزمع, غزة: جامعة القدس المفتوحة, خان يونس.
- الخطايبة, ساميا (2002). تحليل أسئلة شهادة الدراسة الثانوية العامة لمبحث التاريخ في الأردن في ضوء تصنيف بلوم للمجال المعرفي. رسالة ماجستير غير منشورة، جامعة اليرموك، الأردن.
- الخوالدة، ناصر والمشاعلة، مجدي والقضاة، محمد (2007). دراسة تقويمية لأسئلة امتحانات شهادة الدراسة الثانوية العامة الأردنية في مبحث العلوم الإسلامية للأعوام (1997-2005) في ضوء المستويات المعرفية. مجلة جامعة النجاح للأبحاث, 21(2)، .
- ساعد, صباح وعامر, وسيلة (2017). تقييم كفاية بناء الاختبارات التحصيلية لدى أساتذة التعليم الجامعي وفق معايير الاختبار الجيد. مجلة العلوم الإنسانية والإجتماعية, (28), جامعة محمد خيضر, الجزائر, بسكرة.
- سايحي, سليمة (2012). قلق الامتحان وبعض العوامل المساعدة لظهوره لدى التلاميذ. مجلة العلوم الإنسانية والإجتماعية, (7), الجزائر: جامعة محمد خيضر بسكرة.
- سنان, إيناس (2006). درجة معرفة وممارسة معلمات الرياضيات في المرحلة المتوسطة لمهارات بناء الاختبارات التحصيلية بمدينة مكة المكرمة. رسالة ماجستير غير منشورة، جامعة أم القرى، كلية التربية، المملكة العربية السعودية.

- سيف، محمد (2008). مستوى معرفة المشرفين التربـويين بمبـادئ بنـاء الاختبـارات التحصيلية فى محافظة تعز. رسالة ماجستير غير منشورة، كليـة التربية، جامعـة اليرمـوك، الأردن.
- شبانة، محمود (2000). تقييم الاختبارات التحصيلية في مادة الرياضيات لطلبة الصف العاشر الأساسيين في منطقة لواء وادي السير. رسالة ماجستير غير منشورة، جامعة القديس يوسف، بيروت.
- عدائكة، دنيا وجخراب، محمد وجعفور، ربيعة (2018). مستوى معرفة أساتذة التعليم الثانوي لكفايات إعداد الاختبارات التحصيلية الجيدة في ضوء أدوارهم الجديدة(دراسة استكشافية بثانويات ولاية الوادي وسط). مجلة الباحث في العلوم الانسانية والاجتماعية، (33)، 1072-1055.
- علوان, عبد الله (2006). دراسة تحليلية لأسئلة كتابي (المطالعة والأدب والنقد) للصف الحادي عشر
- الفرجات، هشام (2015). مدى كفاءة معلمي المدارس الحكومية الثانوية التابعة لمديرية التربية والتعليم بمحافظة العقبة في بناء الاختبارات التحصيلية فق معايير الاختبار الجيد. مجلة كلية التربية، جامعة الأزهر، 164(2)، 600-565.
- القحطاني، سالم (1996). 69-93). تقويم أسئلة الاختبارات النهائية والأسئلة الواردة في كتب الدراسات الاجتماعية للمرحلة المتوسطة في منطقة أبها التعليمية. المجلة التربوية,41(11) ، 93-69، جامعة الكويت.
- الكحلوت، إيمان (2004). تقويم اختبارات اللغة العربية لصف الثالث الأساسي في ضوء معايير الاختبار الجيد. رسالة ماجستير غير منشورة، كلية التربية، الجامعة الإسلامية، غزة.
- المطاوعة, فاطمة (2000). 27-57). أسئلة القراءة وامتحاناتها في المرحلة الابتدائية بدولة قطر. مجلة البحوث التربوية, 18(9)، 57-27، جامعة قطر.
- الهياجنة, أحمد (1998). تقويم الأسئلة الكتابية الاختبارية التي يعدها معلمو اللغة العربية في المرحلة الثانوية وفق تصنيف بلوم. رسالة ماجستير غير منشورة, جامعة اليرموك, الأردن.
- وزارة التربية والتعليم (2000). مرشد المعلم في بناء الاختبارات التحصيلية. عمان, الأردن: منشورات وزارة التربية والتعليم.

المراجع الأجنبية:

- Anastasi, A & Urbina, S (1997). *Psychological Testing*. 7th Ed., New York: Prentice Hall.
- Angoff, W. (1982). *Norms and scales.* New York: MaMillan.
- Angoff, W. (1971). *Scales, norms, and equivalent scores.* Washington: American Council on Education.
- Bean, J. (1993). The Application of a Model of Turnover in Work Organizations to the Student Attrition Process**.** *Review of Higher Education,* 6(2), 129-148.Black, R. (2018). An analysis of levels of thinking in Nigerian Science teachers examination. *Journal of Research in Science Teaching***,** 17(4), 301-3 06
- Cassidy, S. (2009). Interpretation of competence in student assessment. *Nursing Standard,* 23(18), 39-46.
- Cronbach, L. (1997). *Essential of psychological testing*. 3rd ed, New York: Harper & Row.

- Domino, G. & Domino, M. (2006). *Psychological Testing.* (2nd ed.). Washington: John Wiley & Sons.
- Ebel, R. (2006). Essentials of educational measurement. 8th ed, New Jersey: Engle Wood Cliffs, Prentice-Hall, Ic,
- Flanagan, J. (1993). *Units, scores and norms.* New york: Henry Holt Co. Inc.
- Fleming, M. & Chambers, B. (2015). Teacher made tests, windows on the classroom. *(ERIC Document Reproduction Service no. Ed 182465).*
- Francis, H. (2018). The Influence of Analysis and Evaluation Questions on Achievement in Sixth Grade. *Social Educational leadership Research Supplement,* 5(23), 328-331.
- Friedlend, A. & Folt, C. (2015). *Writing successful science propsals*. New Haven, CT, USA: Yale University Press.
- Hambleton, R. ; Merenda, P. & Spielberger, C. (2005). *Adapting education and psychological tests for cross-cultural assessment.* New York: Lawrene Trlbaum Association Inc.
- Hinter, J. (2017). Qualitative analysis of seven general education progress tests for high level thinking skills in Grade 4 and 6. *DissertationAbstracts International***.** 54 (8), 2867 A.
- Hodges, B. (2009). Factors that can influence mentorship relationship. *Nursing Standard*, 21(6), 33-35.
- Holland, P. & Rubin, D. (2009). *Test equating.* 6th ed, New York: Princetion, Academic Press.
- Gregory, R. (2011***).*** *Psychological Testing: history, principles, and applications.* (6th. ed.). London: London: Rinehart and Holt.
- Kaplan, R. & Saccuzzo, D. (2009). *Psychological Testing: Principles, Applications, and Issues.* 7th. ed. Belmont, USA: Wadsworth.
- Kelley, T. (2001). *Tests and measerements in the social sciences.* 9th ed, New York: Scribner.
- Moulton, M. (2013). Effects of level review and sequence of inserted questions of test processing . *Journal of educational psychology*,73,427- 436.
- Pfeiffer, 1. & Davis, O. (2016). Teacher-made examinations, What kind of thinking do they demand? *National Association of Secondary School Principles***,** 49, 1-10.
- Risner. G. (2017). *Cognitirellerels of Questions Demonstrated by Test Items That Accompany Selected Fifth Grade Science text books***.** Diss .George Peabody College of Teachers.
- Russell, K. (2001). *Psychometric Testing***.** Washington: John Wiley & Sons.
- Schmit, N. & Drasgow, F. (2002). *Measuring and Analyzing Behavior in Organizations: Advances in Measurement*. Boston: Bass-Jossey.
- Schneidler, G. (2007). *Grade and age norms for the Minnesota vocational test.* New Jersey: World Book Com.

- Serkan, Y. (2013). *Analysis Of Turkish High- School Chemistry-Examination Questions According To Blooms Taxonomy*. Chemistry Research And Practice.