أثر طول الاختبار على قدرة الفرد وصعوبة المفردة فى ضوء نظرية الاستجابة للمفردة

إعـــداد

أ.د/ محمد أحمد إبراهيم غنيم

أستاذ علم النفس التربوى ورئيس مجلس قسم علم النفس التربوى بكلية التربية جامعة بنها

أ.د/ مسعد ربيع عبد الله أبوالعلا

أستاذ علم النفس التربوى بكلية التربية - جامعة بنها

أ/ إيمان عبد الفتاح السيد

مدرس مساعد بقسم علم النفس التربوى

بحث مشتق من الرسالة الخاصة بالباحثة

أثر طول الاختبار على قدرة الفرد وصعوبة المفردة فى ضوء نظرية الاستجابة للمفردة

إعـــداد

أ.د/ محمد أحمد إبراهيم غنيم
أستاذ علم النفس التربوى ورئيس مجلس قسم علم النفس التربوى بكلية التربية جامعة بنها

أ.د/ مسعد ربيع عبد الله أبوالعلا
أستاذ علم النفس التربوى بكلية التربية - جامعة بنها

أ/ إيمان عبد الفتاح السيد
مدرس مساعد بقسم علم النفس التربوى

مقـــدمـــة الـــدراســـــة:

تعد الاختبارات التحصيلية من أهم الأدوات التى تزودنا بالبيانات التى يُستند إليها فى اتخاذ العديد من القرارات التربوية التى تخص الفرد والمجتمع، حيث تُصمم لأهداف متنوعه منها اختيار شخص لوظيفة ما أولأغراض التصنيف لتحديد مسار الطلاب بما يتناسب مع قدراتهم ومهاراتهم من خلال تقويم تحصيل الطلاب فى ضوء الدرجات التى يحصلون عليها فى الاختبارات الصفية.

كما يعد التطور الذى حدث فى الاختبارات والعملية الاختبارية أساساً لتحسين العملية التدريسية وذلك لأن القياس ضرورى لتطوير التعليم وتعد العملية الاختبارية ذات دور بالغ الأهمية للعمل التربوى والعملية التدريسية بخاصة وليست منفصلة عن أى منهما فعلى المعلم أن يدرك أن أهداف العملية الأختبارية يجب أن تتطابق مع أهداف العملية التدريسية مما يتطلب تطوير أساليب اختبارية متنوعة بدرجة كبيرة وكل من هذه الأساليب يكون له إسهاماته المتفردة فى عملية القياس (صلاح علام، ٢٠٠٧: ٢٧).

والجدير بالذكر أن الاختبارات التحصيلية التى يُعتنى ببنائها يمكن أن تستثير دافعية الطلاب للتحصيل الجيد وتحقيق الأهداف التعليمية بمختلف مستوياتها وهذا يساعد فى استبقاء المعلومات وتوظيفها وانتقال أثر تعلمها إلى مواقف خارج حدود المنهج المدرسي، كما يساعد فى تأكيد أهمية التقويم الذاتى الذى يعد مكونه أساسية من مكونات التعلم المستقل للطالب وزيادة دافعيتة وتقديره لذاته وبذلك يعتمد الطالب على الضبط والتحكم غير المباشر والدافعية الداخلية فى تحقيق الاهداف التعليمية المرجوة (صلاح علام، ٢٠٠٦: ١٣١)٠

ويذكر (Kesamang, 2017, 3) أن المفردة تخلق نوع من السلوك لدى الممتحن، والممتحن يكتسب نوع من المقدرة (θ) المعرفية للتغلب على المفردة، كما تتمتع المفردة بمقاومة معرفية داخلية (b) لذا فإن إمكانية تغلب الممتحن على المفردة تعتمد على هذين العاملين (b, θ)،

إذا كانت قدرة الممتحن المعرفية أعلى من المقاومة المعرفية للمفردة إذاً فإن احتمالية أن يتغلب الممتحن على المفردة تكون أعلى، لذلك نجد أن القياس الجيد يتطلب:

- أن يتوفر للشخص الذي يتمتع بمقدرة عالية فرصة أفضل للتغلب على المفردة من الشخص ذو المقدرة الضعيفة.
- أن يتمتع أي شخص بفرصة أفضل للتغلب على المفردة ذات المستوى الاقل less demanding item عن المفردة الأكثر تطلباً.
- أن تكون هذه الظروف فقط بمثابة تداعيات لموقع الشخص والمفردة على السمة الخاضعة للقياس.

وتشير (أمينة كاظم، ١٩٩٦: ٤٣١) إلى أن استخدام نموذج راش فى بناء الاختبارات التحصيلية وتكوين بنوك الاسئلة يمكنا من التغلب على مشكلتين من أهم مشكلات التقويم الاكاديمى وهما:

- موضوعية تقدير المستوى الأكاديمى للطالب بالرغم من اختلاف الاختبارات المستخدمة وتباين مستوياتها.
- موضوعية تقدير المستوى الاكاديمى للطالب بالرغم من اختلاف مستوى الشعبة أو المجموعة التى ينتمى إليه.

ويُعد الأختلاف الرئيسى بين النماذج الكلاسيكية ونماذج السمات الكامنه فى مجال تصميم وبناء الاختبارات يبرز أثناء تحليل مفردات الاختبار وانتقاء مفرداته وتقدير ثبات البيانات المستمدة منه والتى تتميز بخاصية عدم التباين Invariance وهوما لا يحدث عند استخدام النماذج الكلاسيكية، فبناء الاختبارات يمر بعدة خطوات من أهمها: إعداد مواصفات الاختبار، وكتابة مفرداته، وتحديد الخصائص السيكومترية لهذه المفردات وانتقاء المفردات الصالحة، وإعداد معايير الاختبار إذا كان جماعى المرجع أوتحديد درجات القطع إذا كان محكى المرجع، ثم تقدير ثبات البيانات المستمدة منه والتحقق من صدق الاختبار، وتختلف نماذج السمات الكامنه عن النماذج الكلاسيكية فى الخطوات المتعلق بتحديد خصائص المفردات وانتقاء المفردات الصالحة وتقدير الثبات (صلاح علام ،١٩٨٧: ٣٥-٣٦).

وعلى الرغم من تعدد نماذج نظرية الاستجابة للمفردة واختلافها فى عدد بارامتراتها وأغراض استخداماتها إلا أن كثير من الدراسات وخبرات الدول المتطورة أكدت أن نموذج راش يُعد أفضل هذه النماذج فى بناء بنوك الاسئلة (صلاح علام، ٢٠٠٥: ٢٧١).

مشكلـــة الـــدراســـــة:

أُجريت على الاختبارات التحصيليه العديد من الدراسات بهدف التعرف على العوامل المؤثرة على بارامتري نظرية الأستجابة للمفردة (قدرة الفرد – صعوبة المفردة) ومنها متغير طول الاختبار ومنها: **دراسة Bastari (2000)**؛ التى هدفة إلي بحث أثر طول الاختبار فى مادة اللغة الانجليزية فى مراحل تعليمية متعددة وحجم العينة ونمط المفردة – ثنائية الدرجة "صواب وخطأ" ومتعددة الدرجة "اسئلة الاجابة القصيرة"– علي تقدير قدرة الفرد باستخدام النموذج اللوغاريتمى ثلاثى المعلم لمفردات الاختيار من متعدد ومفردات الاستجابة المتدرجه –الاستجابة التكوينية–، ولدراسة أثر طول الاختبار أستخدمت الدراسة اختبارين الاول يتكون من (٣٠) مفردة والثانى يتكون من (٥٠) مفردة وقد استخدم برنامج MULTILOG لتقدير معالم المفردات، وأسفرت النتائج عن أنه كلما زاد طول الاختبار كلما زادت دقة معايير المفردة، **دراسة الشيماء عبد الحميد محمد (٢٠٠٧)**، والتى هدفت الى بحث أثر طول الاختبار على دقة نتائج الاختبارات التحصيلية مرجعية المحك فى الرياضيات باستخدام نموذج راش، على عينة من طلاب وطالبات الصف الأول الإعدادى، أعدت الباحثة أداة الدراسة من مفردات تقيس الأهداف السلوكية لمحتوى وحدة فى مادة الجبر وقد بلغ عددها (٩٠) مفردة اختبارية من نوع الصواب والخطأ، وأوضحت نتائج الدراسة أن طول الاختبار يؤثر فى دقة نتائج الاختبارات باستخدام نموذج راش حيث تشير النتائج الى أن أفضل دقة لتقدير القياس تحقق عندما كان طول الاختبار (٩٠) مفردة، وأفضل مطابقة للمفردات مع نموذج راش تتحقق عندما كان طول الاختبار (٦٠) مفردة، ويزداد قدرة الأفراد بزيادة طول الاختبار، وأفضل أتساق داخلى بين المفردات تحقق عندما كان طول الاختبار (٩٠) مفردة، وحقق طول الاختبار (٣٠) مفردة أفضل تمييزاً بين الأفراد ،وبالتالى فإن نتائج الدراسة تشير إلى أن أفضل دقة لتقدير القياس تحقق بزيادة طول الاختبار وكانت أطوال الاختبارات هى (٣٠ – ٦٠ – ٩٠) مفردة، وأفضل مطابقة للمفردات مع نموذج راش تتحقق عند الاختبار متوسط الطول، ويزداد قدرة الأفراد بزيادة طول الاختبار، وأفضل أتساق داخلى بين المفردات تحقق بزيادة طول الاختبار، أفضل تمييزاً بين الأفراد يتحقق بقصر طول الاختبار، دراسة **(Sahin, A. & Weiss, D., 2015)** تهدف هذه الدراسة إلى بحث تأثير بنوك للعينة تمثل بنوك بأحجام صغيرة (١٠٠مفردة) ومتوسطة (٢٠٠ مفردة و٣٠٠ مفردة) من بنك (٥٠٠) مفردة، خضعت المفردات في هذه البنوك لإعادة

التدريج باستخدام العينات المختارة مع استخدام بارامتراتها المقدرة في المحاكاة البصرية الخاصة من أجل إعادة تقدير بارمترات القدرة للعشرة آلاف ممتحن المختارين، أظهرت الدراسات أنه يمكن الحصول على تقديرات دقيقة لمؤشر القدرة مع عينة تدريج من ١٥٠ ممتحن، علاوةً على ذلك يمكن استخدام بنك من ٢٠٠ مفردة قبل التدريج مع ١٥٠ ممتحن من أجل بعض الأهداف في الاختبار التواؤمي باستخدام الكمبيوتر طالما أنه يتمتع بالمعلومات الكافية على مستويات القدرة المستهدفة، تسعى الدراسة للاجابة على التساؤل التالى كيف تتمكن تقديرات القدرة θ بناءً على تقديرات بارمتر المفردة المستقاة من أحجام متنوعة للعينة والبنك من استعادة تقديرات القدرة θ من بنك ضخم للمفردات خضع للتدريج باستخدام عينة ضخمة جداً من الممتحنين؟، تمت محاكاة مجموعة البيانات الكاملة لهذه الدراسة باستخدام 3PLM وإجراء مونت كارلوللمحاكاة في نسخة برمجيات الاختبار التواؤمي الحاسوبي 6 , .4, مع توزيع موحد لبارامترات θ بين (٣- و٣+) 3PLM، كما تمت الدراسة على ٤ مجموعات بيانية مع ٥٠٠، ٣٠٠، ٢٠٠ و١٠٠ مفردة واستجابات ١٠ آلاف ممتحن، جاء رسم عينات الممتحنين المستخدمة في إعادة تدريج المفردات في بنوك ١٠٠، ٢٠٠ و٣٠٠ مفردة بناءً على مجموعات البيانات التي تضم الاستجابات الأصلية (المحاكاة) للممتحنين إزاء المفردات المختارة، كما تم تكوين عينات التدريج حتى ينعكس توزيع القدرة θ للممتحن في المجموعة الكاملة للبيانات في العينات المختارة، وتوصلت نتائج الدراسة إلى أن معاملات الارتباط بين تقديرات القدرة θ – التي حصلنا عليها مع بنوك ١٠٠، ٢٠٠، ٣٠٠، ٥٠٠ مفردة– وأحجام متنوعة للعينة – تتراوح بين ١٥٠ وحتى ٥٠٠٠ – كانت أعلى من. ٩٤٠ و٠,٩٨. مثل هذه المعاملات المرتفعة تشير إلى علاقات خطية إيجابية قوية بين θ الحقيقية والتقديرية بغض النظر عن حجم البنك أوعدد الممتحنين، كما بدت الفروق في متوسطات الجذر التربيعي لتقديرات قدرة الأفراد في أدنى قيمة لها بين بنوك ٨٠٠، ٣٠٠، ٥٠٠ مفردة بالنسبة لأحجام العينة من ٢٠٠٠ أوأقل وزادت قليلاً لأكبر حجمين للعينة، ولكن الفروق بين المعاملات كانت بسيطة بين بنوك ٣٠٠، ٢٠٠، ١٠٠ مفردة، على سبيل المثال وصلت معاملات الارتباط التي حصلنا عليها من بنوك ٣٠٠، ٢٠٠، ١٠٠ مفردة مع ٢٥٠ ممتحن إلى. ٦١٤.، .٦١٤ و٠,٥٦٧. ولكن نجد الارتباط الناتج عن بنك ٥٠٠ مفردة في نفس الظروف. ٧٤٨. وبالتالى فإن نتائج الدراسة تشير إلى أن طول بنك المفردة يحدث تأثير إيجابي على تقديرات θ في هذا المدى حيث أن المزيد من المفردات في بنك المفردة ينتج عنه مزيد من المعلومات، ولأن مستوى معلومات البنك يعتمد بشكل كبير على

عدد المفردات فيه، نجد أن البنك الذي يضم ٢٠٠ مفردة أوأكثر قد يعمل بشكل أفضل لخدمة معظم الأهداف المرجوة وقد يقلل من خطر الحصول على تقديرات θ غير دقيقة في معظم المواقف، علاوةً على ذلك في معظم المواقف نجد أن بنك ٣٠٠ مفردة يعمل بنفس كفاءة بنك٥٠٠ مفردة قد يعني ذلك أنه عند التخطيط لتطوير بنك مفردة ضخم يجب التفكير في بنك ٣٠٠ مفردة، تتسق نتائج هذه الدراسة جزئياً مع أدبيات البحث السابقة حيث يوجد عدد محدود من الدراسات المماثلة (بنك من ٢٠٠ مفردة مدرج على ٢٠٠٠ ممتحن) هوما تم اقتراحه بوصفه ضرورة للحصول على تقديرات θ دقيقة باستخدام الاختبار التواؤمي في الدراسة الحالية وجدنا أن بنك ٢٠٠ مفردة يتمتع بجدوى مناسبة لبعض الاهداف، وبالمثل ظهرت جدوى عينة تدريج من ١٥٠ ممتحن للحصول على تقديرات θ جيدة في مواقف يتمتع فيها البنك بالمعلومات الكافية من أجل θ المستهدفة، **دراسة (Sahin & Aml (2016**، والتى هدفت إلى دراسة أثر طول الاختبار على بارامترات المفردة (الصعوبة، التمييز، التخمين) فى نظرية الاستجابة للمفردة باستخدام نماذج ثنائية التفرع أحادية البعد لنظرية الاستجابة للمفردة (unidimensional dichotomous models)، يشارك في هذه الدراسة ٦,٢٨٨ طالباً جديداً في فصل (دورة) اللغة الإنجليزية في جامعة كبيرة، أما أداة جمع البيانات فهي عبارة عن اختبار لغوي فى الغة إنجليزية مكون من ٥٠ مفردة ولتحقيق متطلبات صدق المحتوى، يتم تدوين المفردات في الاختبار بشكل متوازي مع أهداف الدورة على يد معلمي اللغة الإنجليزية الذين يدرسون الدورة ويعتبرون خبراء في المادة الدراسية، علاوةً على ذلك قبل تقديم الاختبار تتلقى الجامعة التي تشهد جمع البيانات آراء الخبراء من مرشح لنيل درجة الدكتوراه ومتخصصين في اختبارات اللغة يعملون في مكتب الاختبارات، بهذا تخضع المفردات للمراجعة والتنقيح عدة مرات بناءً على آراء الخبراء حتى الوصول لشكلها النهائي، من ثم يتم تقديم الأداة بشكل متزامن إلى ٦,٢٨٨ شخص هم عينة الدراسة في جلسة واحدة، استخدمت البيانات المستقاة من هذا الاختبار للحصول على مجموعات بيانية من ثلاث أطوال مختلفة للاختبار (١٠، ٢٠، ٣٠ مفردة)، من ثم يتم استخدام هذه المجموعات البيانية لخلق ظروف بحثية متنوعة حيث يتم التلاعب في طول الاختباروحجم العينة ومتغيرات نموذج نظرية الاستجابة للمفردة وذلك لدراسة مدى دقة تقدير برامترات المفردة في ظل ظروف مختلفة، أظهرت النتائج المستقاة هنا أنه بجانب حجم العينة أوطول الاختبار يُعد المزيج المكون من هذين المتغيرين هام للغاية ويمكن الاستعانة بالعينات التي تتكون من

١٥٠، ٢٥٠، ٣٥٠، ٥٠٠ و٧٥٠ مفحوص لتقدير بارامترات المفردة بدقة في ثلاث نماذج ثنائية التفرع أحادية البعد لنظرية الاستجابة للمفردة بالاعتماد على طول الاختبار والنموذج المستخدم في هذه الدراسة خضع للاختبار ٩ من أحجام العينات الصغيرة والكبيرة في نفس الوقت وهوما يجعل من الممكن مقارنة ومقابلة النتائج المستقاة من الجانبين، وهذا ما يجعل هذه الدراسة فريدة من نوعها في عالم أدبيات البحث في ظل هاتين السمتين، تعمل هذه الدراسة على تزويد كل ممارسي نظرية الاستجابة للمفردة بمخطط تفصيلي يشرح كيف يكون حجم العينة صغير أوكبير لتقديرات بارامترات المفردة بدقة عند وضع اختبار يقوم على نظرية الاستجابة للمفردة عن طريق الإجابة على السؤال التالي: "كيف يؤثر حجم العينة وطول الاختبار على تقدير مؤشر المفردة في تطوير الاختبار القائم على نظرية الاستجابة للمفردة؟" وبالتالى فإن نتائج الدراسة تشير إلى ضرورة وضع في الاعتبار الدراسات السابقة في أدبيات البحث عند اختيار المفردات للاختبارات القصيرة من المجموعة الكاملة للبيانات لصياغة اختبارات فرعية ذات أطوال متنوعة وأثناء تحديد عدد المفردات في الاختبارات القصيرة، **دراسة (Jiang, S.et.all, 2016)** أجريت دراسة محاكاة لاستعراض المتغيرات التي قد تؤثر في تقدير بارامتر المفردة من أجل نموذج الاستجابة المدرجة متعدد الأبعاد، تم استخلاص البيانات بناءاً على أطوال متنوعة للاختبار في نموذج الاستجابة المدرجة أحادي البعد، حصلت الدراسة على تقديرات البارامتر عبر برمجيات فليكس MIRT، خضعت جودة تقدير البارامتر للتقييم عن طريق الارتباط بين المؤشرات الحقيقية والتقديرية وكذلك التحيز وخطأ جذر متوسط التربيع، خضعت ثلاث متغيرات للمعالجة هى: (١) طول الاختبار:L = ٣٠ و٩٠ و٢٤٠ مفردة، استخدمت الدراسة التي بين أيدينا نموذج الاستجابة المدرجة ثلاثي الأبعاد مع ٤ تصنيفات للاستجابة لكل المفردات، نتج عن الاختبار الأقصر خطأ أقل في جذر متوسط التربيع لمؤشرات الحد، ولكن لمؤشرات التمايز لم يظهر أي تغيير في مؤشرات الحد مع تزايد طول الاختبار من L= ٣٠ وحتى L= ٩٠، ولكن زاد الخطأ في جذر متوسط التربيع لتقديرات صعوبة المفردة من L= ٩٠ وحتى L= ٢٤٠ بمقدار حوالي. ٠٤. للحدين وحوالي. ٠٢. للحد المتوسط وبالتالى فإن نتائج الدراسة تشير إلى أن مقدار جذر متوسط مربع تقديرات صعوبة المفردة يزاد من ٩٠ مفردة إلى ٢٤٠ أى أن الأفضلية للاختبار القصير ٣٠ مفردة فى تقدير بارامتر صعوبة المفردة، **دراسة (Hidalgo, M.et.all, 2016)** تهدف إلى مقارنة التراجع اللوجستي المميز واختبارات نسبة الاحتمال لنظرية الاستجابة للمفردة من أجل الدالة المميزة للمفردة في الاختبارات

(IRT/ RDIF)، اتبعت الدراسة فى معالجة استجابات المفردة احصائياً نموذج الاستجابة المتدرجة، من المتغيرات التى تم معالجتها طول الاختبارجيث تتألف الاختبارات المحاكاة من ٤، ٥، ٨ أو ١٠ مفردات المقياس الذي يتألف من ١٠ مفردات، توصلت نتائج الدراسة إلى أن معدل الخطأ فى تقدير بارامترات النموذج يتناقص بزيادة طول الاختبار.

ويتضح من نتائج الدراسات السابقة:

- الاختلاف بين نتائج الدرسات والأبحاث حول أفضلية أطوال الاختبارات المستخدمة فى التحليل، حيث أتفقت نتائج دراسة كل من: **Bastari(2000)، الشيماء عبد الحميد محمد (٢٠٠٧)، Sahin& Weiss (2015)، Jiang, et all(2016)، Hidalgo, et all (2016)** على أنه بزيادة طول الاختبار يزداد دقة بارامترات المفردات – صعوبة المفردة وقدرة الفرد – المكونة للاختبار، بينما أوضحت نتيجة دراسة: **Sahin&Aml(2016)**، أن الاختبار القصير يعطى تقديرات لصعوبة المفردة أفضل من الأختبارات الأطول.
- لا يوجد فى حدود علم الباحثة درسات عربية تناولت أهدافها متغير طول الاختبار على اختبار تحصيلى موضوعى فى مادة العلوم للصف الثانى الأعدادى لبحث أثر متغير طول الاختبار على قدرة الفرد وصعوبة المفردة.

وبذلك يمكن تحديد مشكلة الدراسة فى التساؤل التالي:

هل تختلف تقديرات معلم صعوبة المفردة ومعلم قدرة الفرد باختلاف طول الاختبار الموضوعى المستخدم (٢٨ – ٥٣ – ١٠٠) مفردة؟

أهـــداف الـــدراســـــة:

يهدف البحث الحالى إلى بحث أثر متغير طول الاختبار على قدرة الفرد وصعوبة المفردة وذلك باستخدام نظرية الاستجابة للمفردة.

أهميـــة الـــدراســـــة:

تتمثل أهمية البحث الحالى فى حاجة المؤسسات التربوية إلى وجود أدوات علمية دقيقه تستخدم فى عمليات التشخيص والتصنيف ويُعد الاختبار التحصيلى من أكثر الأدوات المستخدمة فى تقييم التحصيل المدرسى الأمر الذى يُسهل القيام بالمهام التعليميه على نحو أكثر فاعليه، ونظرا لأهمية الاختبار التحصيلى تسعى الدراسة الحالية محاولة التعرف على أثر طول الاختبار على بارامترى المفردات المكونه للاختبار التحصيلى؟

وبالنظر الى واقع مؤسساتنا التعليمية نلاحظ قلة وندرة استخدام نظرية الاستجابة للمفردة كأسلوب احصائى فى التعامل مع نتائج الاختبارات التحصيلية للاستفادة من فوائدها فى العملية التعليمية مع عدم وجود توجيه جاد للاهتمام بعملية القياس وتصميم الاختبارات التحصيلية ذات موضوعية فى القياس حسب ما توفرة نماذج نظرية الاستجابة للمفردة، وهوما يحتم على المهتمين بالنواحى التربوية الاهتمام بذلك الجانب الحيوى والهام.

مصطلحات الدراسة:

طول الأختبار:

ويقصد به عدد المفردات الخاصة بكل صورة اختبارية يؤديها الممتحن حيث تختلف الصورتان الاختباريتان فى عدد المفردات التى تتضمنها لكنهما يتماثلان فى قياس نفس الاهداف السلوكية التى تتضمنها الوحدة الدراسية محل اهتمام الباحثة.

معلم صعوبة المفردة:

هى اللوغاريتم الطبيعى لمرجح الخطأ لدى الأفراد الذين تعبر نقطة صفر التدريج عن قدرتهم أى أن صعوبة المفردة دالة لكل من عدد الأفراد الذين أجابوا صواباً على المفردة وكذلك قدرات هؤلاء الأفراد (أنور الشرقاوى، ١٩٩٦:٣٢٤).

معلم قدرة الفرد:

هى اللوغاريتم الطبيعى لمرجح نجاح الفرد على البنود التى تعبر نقطة صفر التدريج عن صعوبتها أى أن قدرة الفرد دالة لكل من عدد المفردات التى أجاب هذا الفرد عليها صواباً وكذلك دالة على صعوبة هذه المفردات(أنور محمد الشرقاوى ،١٩٩٦:٣٢٥).

نظرية الاستجابة للمفردة:

تفترض نظرية الاستجابة للمفردة نماذج رياضية للسمات الكامنة، فهى تفترض أن القيمة الاحتمالية لاستجابة مفحوص لفقرة اختبارية داله لكل من قدرة المفحوص – التى يفترض أن الاختبار يقيسها – وخصائص الفقرة التى يحاول المفحوص الاجابة عنها ومن ثم فإن الدالة المميزه لخصائص المفردة تحدد العلاقة بين مقدار هذا الاحتمال وبين البرامترات المناظرة لقدرة المفحوص وخصائص الفقرة، ووفق لنماذج هذه النظرية فإن تقديرات قدرة المفحوصين متحررة من تقديرات برامترات المفردات كما أن خصائص الفقرة متحررة من خصائص المفحوصين الذين

طبقت عليهم ونظرا لأن استجابة المفحوص للمفردة تتضمن اخطاء فان هذه النظرية تهدف للتوصل لقيم تقديرية للبارامترات ومن ثم استخدام هذه القيم فى تقدير احتمال الاستجابة الصحيحة لكل مفردة من مفردات الاختبار، أى أن أداء الممتحن في الاختبار يمكن التنبؤ به من خلال مجموعة من العوامل تسمى السمات والعلاقة بين أداء الممتحن في الاختبار والسمات التي من المفترض أن تؤثر على الأداء، والأداء على الاختبار يمكن وصفه من خلال دالة متزايدة طردياً تسمى المنحنى المميز للمفردة، ويمثل هذا المنحنى الانحدار غير الخطى للـ (درجات التى يحصل عليها المفحوصون فى كل مفردة من مفردات الاختبار) على (القيم التقديرية للسمة أوالقدرة المقاسة) ومنه يتضح قدرة المفحوص المناظرة لاحتمال معين للأداء على فقرة ما، وهناك خمسة افتراضات شائعة لنظرية الاستجابة للمفردة وهم: (١) أحادية البعد، (٢) الاستقلال الموضعي، (٣) المنحنيات المميزة للمفردة، (٤) لا تغاير بارامترات المفردة والقدرة، (٥) السرعة، (Oji, E. & Charles, I. , 2016 112:114)، (Akour&Al-Omari, 2013, 291;Gleason, 2008, 9).

الاطـــار النظـــرى للـــدراســـة

من منطلق ما تسعى إليه هذه الدراسة وهوبحث أثر طول الاختبار التحصيلى على بارامترى المفردة – صعوبة المفردة وقدرة الفرد –، لذا تجدر الإشارة إلى أنه شهد مجال القياس النفسي والتربوي في الآونة الأخيرة اهتماماً ملحوظاً على أيدي خبراء القياس النفسي والتربوي في كثير من دول العالم وذلك لبناء مفردات اختبارية تساعد على القياس الدقيق لمخرجات التعلم والحصول على تقديرات كمية تتسم بالموضوعية للسمات التي تنطوي عليها استجابات الطلاب لتلك المفردات الاختبارية، وقد تمثل هذا الاهتمام في ثلاثة اتجاهات رئيسة هي: الاتجاه الأول: تمثل في الاهتمام بالأسس التي يتم في ضوئها تصميم وبناء المفردات الاختبارية لقياس مخرجات التعلم، وقد ساعد هذا الاتجاه على ظهور العديد من التقسيمات التربوية التي تفسر التعلم ومخرجاته (Taxonomies) وتوضح كيفية بناء المفردات الاختبارية لقياس مخرجات التعلم المعرفية، والوجدانية، والمهارية. ومن تلك التقسيمات التي تهتم بدراسة المخرجات المعرفية للتعلم: تقسيم بلوم(Bloom's Taxonomy) وتقسيم بنية ناتج التعلم الملحوظ (The Structure of Observed Learning Outcome)، والاتجاه الثاني: تمثل في الاهتمام بالشكل العام للاختبار النفسي والتربوي والعلاقات الداخلية بين المفردات، وقد ساعد هذا الاتجاه

علـى ظهور نظريـة الاختبـارات العنقوديـة (Testlets) لتوضـيح كيفيـة بنـاء اختبـارات عنقوديـة خطيـة أوهرميـة، والاتجـاه الثالـث: تمثـل فـي الاهتمـام بكيفيـة تحليـل المفردات الاختباريـة لتقدير العلاقـة الاحتماليـة بـين أداء الطالـب فـي الاختبـار والسـمات أوالقـدرات الكامنـة وراء هـذا الأداء وتفسيرها، وقد ساعد هذا الاتجاه على ظهور نظرية الاستجابة للمفردة (Item Response Theory) وما انبثق عنها من نماذج متعددة أكثرها شيوعا نماذج راش (Rasch Models) الاحتمالية، وقد برز هذا الاهتمام في مجال القياس والتقويم كرد فعل للنقد الذي تم توجيهه إلي تقسيم بلوم من ناحية، والنقد الذي تم توجيهه إلي نظرية القياس التقليدية (Classical Test Theory) من ناحية أخري، ومن تلك الانتقادات التي وجهت إلي الافتراضات التي قام عليها تقسيم بلوم: أنه لا يتم تمثيل المعرفة هرميا، ويصعب الفصل بين مستوياته الستة، ويوجد تداخل بين تلك المستويات، كما أنه لا يستطيع قياس المعالجات العميقة للتعلم، فقد يعتمد الطالب على المعالجـة السـطحية فـي الإجابـة عـن أسـئلة ذات مسـتويات عليـا فـي تقسـيم بلـوم، كمـا تعـد الانتقادات التـي وجهـت إلـي فلسفة القيـاس النفسـي والتربـوي عامـة، والنظـام المرجعـي الجماعـة خاصـة، الأسس التـي يستند إليهـا هذا النظـام فـي بنـاء الاختبـارات والمقاييس النفسية والتربوية وتحليـل مفرداتهـا دافعـاً نحوالبحـث عـن نظريـة جديـدة فـي القيـاس، إذ يعـاب علـى الاختبـارات والمقاييس المقننة أنها تعتمد على موازنـة أداء المتعلم بأداء أقرانه وفقا للمجموع الكلي لدرجاتهم الاختباريه وسواء أجريت هذه الموازنة على أساس الدرجات الخام أوالدرجات المعيارية بأنواعها المختلفة، فإن هذه الدرجات تخضـع لخصـائص عينـة الأفراد التـي تستخدم فـي اشتقاق معايير الاختبـار، كمـا تخضـع لخصـائص عينـة الأسـئلة التـي يشـتمل عليهـا الاختبـار واخـتلاف هـذه الخصـائص بـاختلاف عينـة الأفـراد المختبـرين وعينـة الأسـئلة يـؤثر بـلا شـك فـي صـدق هـذه الموازنـات، وبـذلك يصـعب تعمـيم نتـائج الاختبـارات أوالاسـتفادة منهـا، فتـدريج صـعوبة أسـئلة الاختبار يتباين بتباين قدرات الأفراد الذين يختبرون بهذه الأسئلة، وقياس قدرات أوسمات الأفراد يتبـاين بتبـاين صـعوبة الأسـئلة، كمـا أن ثبـات درجـات الاختباريتـأثر بتبـاين قـدرات الأفـراد الـذين يجرى عليهم الاختبار ومستوي هذه القدرات (محمد محمد، ٢٠١٠، ٢).

ويذكر كل من هامبليتون وإجنور (Hambleton & Eignor, 1980) أن هنـاك عاملين ينبغى وضعهما فى الاعتبار عند تحديد أسئلة الاختبار وهما:

١- العلاقة بين عدد أسئلة الاختبار والأهمية الخاصة بكفاية معينة:

تتفاوت أهمية الكفايات المراد قياسها فى الاختبار لذا ينبغى أن يختار مصمم الاختبار عند قياسه كفايات متعددة أسئلة تتناسب مع أهمية النطاق المراد قياسه.

٢- العلاقة بين عدد أسئلة الاختبار ومستوى الحد الأدنى المقبول لدرجة ثبات الاختبار:

يستدعى ذلك تحديد قيمة الدقة المطلوبة (Degree of Precision) فى تقدير درجة الطالب فى النطاق وذلك يتم وفق الصيغة التالية:

$$\text{طول الاختبار} = \frac{0.25}{2^{(\text{درجة الدقة المطلوبة})}}$$

حيث أن (0.25) مقدار ثابت، كما أشار كل من هامبليتون وآخرون (Hambleton et.al., 1978) أن الهدف من تطبيق الاختبار هوتصنيف الطلاب إلى فئتين متقنة وغير متقنة لذا فإن تحديد طول الاختبار المناسب الذى يقلل من أخطاء التصنيف Number of Classification Errors يتطلب تطبيق اختبار طويل يقلل من أخطاء القياس، وأضاف كل من هالادين وروبد (Haladen&Roid) أن قيمة أخطاء التصنيف تقل بشكل عام عند زيادة طول الاختبار لكن النقص فى قيمة الخطأ يكون ملحوظ عند الانتقال من اختبار طوله (١٠) مفردات إلى اختبار طوله (٢٠) مفردة وأما النقص فى قيمة الخطأ عند الانتقال من اختبار طوله (٢٠) مفردة إلى اختبار طوله (٣٠-٤٠) مفردة يكون صغير نسباً (فى: خالد شيبان، ١٩٩٩: ٢٤).

ويذكر (رضا عوض، ٢٠٠٤: ٥٢) أن الزيادة المناسبة فى عدد المفردات تؤدى إلى التحسن فى ثبات درجات الاختبار إلى أن يصل هذا التحسن عند نقطة معينة يبدأ بعدها فى الانخفاض، أى أن الزيادة الاضافية فى عدد المفردات بعد الوصول إلى هذه النقطة لا يقابلها زيادة أوتحسن فى الدرجات نتيجة لانخفاض جهد الطلاب وكذلك الزمن المخصص للإجابة، لذا فعلاقة عدد مفردات الاختبار بثبات درجاته هى علاقة طردية ليست خطية وإنما علاقة طردية منحنية بمعنى أن الزيادة المناسبة فى عدد المفردات ترفع قيمة معامل ثبات الاختبار عند حد معين ينخفض بعده مع زيادة عدد المفردات.

ويشير (Akour & Al-omari, 2013.292) إلى أن طول أي اختبار وعدد الممتحنين الذي نحتاج إليهم لتقديم تقييم جيد لمعايير المفردة, يكون من الصعب تحديدهم، فالعينات الكبيرة والاختبارات الطويلة نحتاج إليها لتقديم تقييمات دقيقة وخاصة عندما يتم تقييم كلا من معايير

المفردة والقدرة ولهذا السبب توجد معظم تطبيقات IRT في مواقف الاختبار كبيرة المقياس، وعلى الرغم من أن العديد من الاختبارات تتألف من عدد صغير نسبيا من الفقرات التي تُخصص لعينات صغيرة نسبيا من الممتحنين (الاختبارات التي يضعها المدرسين) تم البحث على IRT بعينات صغيرة واختبارات قصيرة لتساعد في وضع إرشادات لأطوال الاختبارات وأحجام العينة المحتاج اليها للحصول على تقييمات معيار ثابتة ودقيقة فبالنسبة للنموذج اللوجستي ثلاثي المعيار (3-PL), اقترح Lord١٩٦٨ أن العينات بأحجام أكبر من ١٠٠٠ ممتحن وأطوال اختبار أطول من ٥٠ مفردة, نحتاج إليها من أجل تقييم مناسب لمعيار التمييز للمفردة. وعلاوة على ذلك, قام كلا من Ree, Jensen(١٩٨٠), بتمثيل اختبارات من ٨٠ مفردة وأحجام عينة من ٢٥٠ و٥٠٠ و١٠٠٠ و٢٠٠ مستخدمين النموذج اللوجستي ثلاثي المعيار. فأنتجت هذه الدراسة أن أحجام العينة الأكبر من ٥٠٠, نحتاج إليها لإنتاج تقييمات أكبر دقة من حيث الصعوبة ومعايير التمييز على الرغم من أن معيار التخمين يظهر أنه يتأثر بالزيادة في حجم العينة وأكدت هذه الدراسة أن التقييمات الدقيقة والثابتة لمعايير التمييز والصعوبة تتطلب أحجام عينات كبيرة في خلال مدى واسع من القدرة، كما أن تقييم التخمين يتطلب عدد كبير من العناصر في المستويات المنخفضة جدا من القدرة، وباستخدام النموذج اللوجستي ثنائي المعيار والنموذج اللوجستي ثلاثي المعيار, قدم Hulin et al في عام ١٩٨٢ دراسة حول تأثير حجم العينة وطول الاختبار على دقة المفردة وتقييمات القدرة. فأحجام العينة ٢٠٠ و٥٠٠ و١٠٠٠ و٢٠٠٠ تم استخدامهم, وأطوال اختبار في ١٥ و٥٠ و٦٠ فقرة، فأشارت النتائج أن أقل أحجام عينة وأطوال الأختبار تعتمد على استجابة النموذج وأهداف الدراسة وبالنسبة للنموذج ثنائي المعيار, أوصت Hulin et al, 1982 أن أقل طول للاختبار يكون ٣٠مفردة وأحجام العينة يتكون من ٥٠٠ فرد فعندما يتم عمل تقييم معايير المفردة والقدرة في نفس الوقت في النموذج اللوجستي ثلاثي المعيار, تطلب أطوال اختبار ٦٠مفردة وأحجام عينة لـ١٠٠٠ فرد لأجل تقييمات عالية الدقة، وعلى الرغم من أن المفاضلات بن حجم العينة وطول الاختبار كانت واضحة في النتائج التى حصلوا عليها فعلى سبيل المثال النتائج المعتمدة على اختبارات من ٦٠ مفردة و١٠٠٠ ممتحن, تمثل هؤلاء المعتمدين على اختبارات من ٣٠ مفردة و٢٠٠٠ ممتحن، كما أنSwaminathan وGifford في عام ١٩٨٣, قاما بدراسة تأثير حجم العينة وطول الاختبار

وتوزيع القدرة على دقة الفقرة وتقييم القدرة. حيث قاموا بتمثيل معلومات مستخدمين النموذج اللوجستي ثلاثي المعيار وأحجام العينة ٥٠ و٢٠٠ و١٠٠٠ ممتحن, وأطوال اختبار ١٠ و١٥ و٢٠مفردة، فأشارت نتائج هذه الدراسة أن أحجام العينة ١٠٠٠ممتحن وطول الاختبار لـ٢٠ مفردة أنتجت تقييمات جيدة جدا للصعوبة ومعايير التخمين وتقييمات جيدة إلى حدا ما لمعيار تمييز الفقرة كما وأشارت العديد من الدراسات السابقة إلى أن العينة التي تكون أكبر من ١٠٠٠ ممتحن والاختبار يتكون من مفرد تتطلب تقييمات عالية الدقة لمعايير ثلاثية المفردة, وهذا عندما يكون النموذج اللوجستي ملائم للمعلومات وعلى الرغم من هذا, لا توجددراسة واحدة من هذه الدراسات تم عملها باستخدام معلومات اختبار حقيقية, بل أنه تم استخدام معلومات افتراضية بدلا من ذلك ويوجد مشكلة في هذا الاتجاه, وهي أنه لا يكون معروفا ما إذا كانت سمات معلومات الاختبار الحقيقي تنعكس في المعلومات الافتراضية، وهذا عندما تكون المعلومات الافتراضية وقيم القدرة مُمثلة دائما من توزيع عادي بمتوسط صفر وانحراف قياسي متحد، فنجد أنه عند استخدام معلومات حقيقية, فإنه يكون غير مضمون أن توزيعات القدرة سوف تكون طبيعية فإذا كانت توزيعات القدرة منحرفة, فهذا ربما ينتج تقييمات معايير للفقرة أقل دقة, عن التقييمات الحاصل عليها من التوزيعات التي تكون طبيعية وبالتالي فإن المعلومات الحقيقية ربما تُنتج نتائج مختلفة عندما تقارن بالمعلومات التمثيلية فيما يتعلق بدقة معايير الفقرة المقدرة وبالإضافة إلى هذا فإن استخدام فقط المعلومات التمثيلية ربما يحدد تعميم النتائج على المعلومات الحقيقية.

فـــرض الـــدراســـــة:

نظراً لتباين نتائج الدراسات والبحوث السابقة فى هذا المجال، فقد صاغت الباحثة فرض البحث على النحوالتالى:

تختلف تقديرات معلم صعوبة المفردة ومعلم قدرة الفرد باختلاف طول الاختبار الموضوعى المستخدم (٢٨ – ٥٣ – ١٠٠) مفردة.

خطـــة وإجـــراءات الـــدراســـة:

١- عينـــة الـــدراســـة:

تم اشتقاق عينة الدراسة الحالية من بين تلاميذ الصف الثانى الأعدادى بمحافظة القليوبية فى العام الدراسى ٢٠١٦ / ٢٠١٧ م.

١-١ عينة الدراسة الاستطلاعية:

تكونت عينة الدراسة الاستطلاعية من (١٠٠) من تلاميذ الصف الثانى الأعدادى بمتوسط عمرى (١٣,٣) سنة وانحراف معيارى (. ٣٩.) سنه بمدارس (أسامه بن زيد – بلال بن رباح – محمد متولى شعراوى) بإدارة (بنها التعليمية) حيث تم تطبيق أداتى الدراسة الحالية فى النصف الاول من العام الدراسي بفاصل زمنى بين تطبيق الاختبارين من اسبوع وذلك للوقوف على:

١- مدى وضوح تعليمات الأداة.

٢- مدى فهم واستيعاب أفراد العينة الاستطلاعية لمفردات الاختبار.

٣- الوقوف على الصعوبات التى تواجه الباحثة فى التطبيق المبدئى والعمل على إزالتها قبل التطبيق النهائى.

٤- الوقوف على الخصائص السيكومترية لأداة الدراسة (ثبات / صدق).

١-٢ عينة الدراسة الأساسية:

تكونت عينة الدراسة الأساسية الأولية من (٥٦٩) تلميذ وتلميذة أُختيروا عشوائياً لتطبيق عليهم الاختبارين (المقالي – الموضوعي) فى الوحدة الأولى بمقرر مادة العلوم للصف الثانى الاعدادى حيث حضر (٤٧٦) تلميذ وتلميذة جلستى الامتحانين (المقالى – الموضوعى)، بينما (٣٥) تلميذ وتلميذة حضروا جلسة الاختبار الموضوعي فقط لذا تم استبعادهم من العينه الأساسية و(٥٨) تلميذ وتلميذة حضروا جلسة الاختبار المقالي فقط لذا تم استبعادهم من العينه الاساسية نظرا لتغيبهم يوم التطبيق وذلك من مدارس (ناصر الأعدادية الحديثه بنين – ٢٥ يناير الأعدادية بنات – محمد متولى شعراوى الأعدادية بنين – بلال بن رباح الاعدادية بنين) بإدارة بنها التعليمية بفاصل زمنى بين تطبيق الاختبارين أسبوع، متوسط عمرهم زمني (١٣,٣) سنة وانحراف معيارى (٠,٣٤) سنة وهى التي استخدمت في إجراء التحليل والتدريج والإجابة على تساؤلات الدراسة وجدول (١) يوضح توزيع أفراد العينة.

جدول (١) توزيع حجم العينة النهائى للتلاميذ الذين أنجزوا جلستى التطبيق

الجنس / المدرسة	ذكور	إناث	المجموع
ناصر الاعدادية بنين	١٠٢	-	١٠٢
٢٥ يناير	-	١١٨	١١٨
محمد متولي شعراوى	١٠٠	-	١٠٠
بلال بن رباح	٥٩	٩٧	١٥٦
المجموع	٢٦١	٢١٥	٤٧٦

٢- أداة الـــــدراســـــة:

إستخدمت الباحثة فى الدراسة الحالية للإجابة على فروض الدراسة اختبار محكى المرجع فى مادة العلوم للفصل الدراسى الأول للمرحلة الثانية من التعليم الأعدادى وتم إعداده وفق الإجراءات الأتية التى استندت الباحثة فيها على عدة مصادر فى هذا المجال وهى (صلاح مراد؛ أمين على سليمان، ٢٠٠٥: ١٤٢)، (صلاح علام ،٢٠٠٧: ٢٦)، (Abd.Razak, et.al., 2012, 2205-2207)، (Cunningham, 1986, 154)، (Yousfi, 2012, 367)، (يحيى علوان، ٢٠٠٧: ٢٤)، (صلاح مراد، أمين سليمان، ٢٠٠٥: ٢١)، (طلال يوسف، ٢٠١٢: ٤١)، (أمينة كاظم، ١٩٩٦: ٤٣١) ،(أحمد الكحلاوى، هند الحموري، ١٩٩٨: ٥٩-٦٠)، (Baker, 2001, 157)، (هشام محمد، ١٩٩٩: ١٤٢-١٥٠)، (محمد الشافعى، ١٩٩٦: ٢٠٣-٢٣٠)، (نضال الشريفين ٢٠٠٦: ٨٧-٨٨)، (أعتدال غازى عبد الله، ٢٠٠٠: ١٣٦:١١٥):

١- تحليل دقيق للوحدة الأولى من فى مقرر الفصل الدراسى الأول من المرحلة الثانية من التعليم الأساسى فى مادة العلوم.

٢- كتابة الأهداف السلوكية المحددة المنبثقة عن الأهداف العامة.

٣- عرض تحليل محتوى مادة التدريس على المحكمين المختصين وذلك للحكم على مدى توافر البنود التالية:

- شمولية الاهداف العامة.
- الدقة فى تحليل المحتوى.
- الدقة فى تجزئة الاهداف العامة إلى أهداف سلوكية قصيرة المدى.
- الدقة فى صياغة الأهداف السلوكية.
- محاولة صياغة اختبار من نمط المفردة ذات الاستجابة الثنائية لقياس كل هدف من الاهداف السلوكية.

٤- عقد لقاءات مع المحكمين ومناقشتهم فى ملاحظتهم واستجاباتهم على البنود السابقة التى أعتمدوا عليها فى الحكم على صلاحية مفردات الاختبار.

٥- إجراءات التعديلات التى تم الاتفاق عليها مع المحكمين.

٦- تدريج الاختبار باستخدام برامج الحاسوب الخاصة بنموذج راش الذى ينتمى لنظرية الاستجابة للمفردة.

وصف الاختبار التحصيلى فى صورته الاولية:

قامت الباحثة بطبع الاختبار الموضوعى والذى يتكون من (١٠٠) مفردة من نوع الاختيار من متعدد للتطبيق المبدئ بهدف التأكد من صلاحيته للتطبيق النهائى وتحديد زمن الاجابه الملائم.

حساب صدق الاختبار الموضوعي:

الخصائص السيكومترية لحساب صدق الاختبار:

١- صـــدق المحكمـــين:

تم تقدير صدق المحكمين للوقوف على صدق الاختبار؛ وذلك بعرض الاختبار على مجموعة من السادة المحكمين لأخذ آرائهم من حيث:

١- كفاية التعليمات المقدمة للتلاميذ للإجابة بطريقة صحيحة على الإختبار.

- أسفر ذلك عن وضوح التعليمات بالأتفاق مع أراء المحكمين.

٢- صلاحية المفردات علمياً، ولغوياً.

- أسفر ذلك عن إضافة للسؤال الاول ليكون صحيح علمياً ولذلك تم تعديله من الصيغة (١- تحتوى نواة الذرة على ..... أ- بروتونات والكترونات، ب- الكترونات سالبة، ج- نيترونات متعادلة، د- بروتونات موجبة) ليصبح (١- تحتوى نواة الذرة على ..... أ- بروتونات ونيترونات، ب- الكترونات سالبة فقط، ج- نيترونات متعادلة فقط، د- بروتونات موجبة فقط).
- وتم تعديل السؤال الثانى من الصيغة (تدور الألكترونات حول نواة ذرة العنصر فى مستويات طاقة عددها..... أ- ٢، ب- ٣ ،ج-٧، د-٥) ليصبح (تدور الألكترونات حول نواة ذرة الصوديوم فى مستويات طاقة عددها..... أ- اثنان، ب- ثلاثه، ج-سبعة، د-خمسة).
- كما تم تعديل السؤال الثامن من (أثناء التفاعل الكيميائى تفقد ذرة البوتاسيوم الكترون التكافؤ لأنها ..... ألكترون. أ- أيون سالب، ب-ذرة متعادلة، ج- عنصر شبه فلز، د- أيون موجب) إلى (أثناء التفاعل الكيميائى تفقد ذرة البوتاسيوم الكترون التكافؤ متحوله إلي ..... أ- أيون سالب، ب-ذرة متعادلة، ج- عنصر شبه فلز، د- أيون موجب).

- وتعديل السؤال الثانى عشر من (يستخدم لحفظ قرنية العين٠٠٠٠٠٠أ-النيتروجين المسال، ب- الهيدروجين، ج-الماء) إلى (يقوم الأطباء بحفظ قرنية العين فى مواد مبرده مثل ..... أ- النيتروجين المسال، ب- الهيدروجين، ج-الماء).
- وأيضاً السؤال الحادى والعشرين من (يتجمد الماء عند درجة..... درجة مئويه. أ-٤، ب- ١٠٠، ج-٩٠، د- صفر) إلى (أعلى قيمة لكثافة الماء عند درجة ..... درجة مئويه. أ-٤، ب- ١٠٠،، ج-٩٠، د- صفر).

٣- مناسبة المفردات لتلاميذ الصف الثانى الإعدادى.

- وفقاً لأراء المحكمين تم تعديل السؤال السادس والاربعين من (العدد الذرى لعنصر النحاس ٢٩ لذا فان عدد الالكترونات التى يحتويها مستوى طاقته الاخير هو..... أ-٢، ب-٨ ،ج-١، د-١١) إلى (العدد الذرى لعنصر الصوديوم ١١ لذا فان عدد الالكترونات التى يحتويها مستوى طاقته الاخير هو..... أ-٢، ب-٨ ،ج-١، د-١١) حيث أقتصر المقرر على توزيع الألكترونات فى مستويات الطاقة بحد أقصى ٢٠ ألكترون.

٤- مناسبة كل سؤال للمستوى الذى وضع لقياسه.

- تم تعديل السؤال الحادى والثلاثين من الصيغة (يذوب ملح الطعام فى الماء لأن الماء٠٠٠٠ أ- مذيب تساهمى، ب-مذيب قطبى ضعيف، ج-مذيب قطبى جيد د-مذيب أيونى)٠ إلى (يذوب $NaCl_2$ فى H_2O لأن الماء .... أ- مذيب تساهمى، ب-مذيب قطبى ضعيف، ج-مذيب قطبى جيد د-مذيب أيونى) ليتناسب مع مستوى التطبيق.
- كما تم تعديل السؤال السادس والثلاثين من (يوصل الماء النقى للتيار الكهربى عند إضافة قطرات من.. ٠٠٠أ- سكر الطعام، ب-كلوريد الهيدروجين، ج- البنزين، د- حمض الكبرتيك)إلى (عند تحليل H_2O كهربياً نضيف..... إلى H_2O النقى أ- سكر الطعام، ب-كلوريد الهيدروجين، ج- البنزين، د- حمض الكبرتيك) ليتلاءم مع مستوى التطبيق.

٣- الصـــدق التمـــييزى:

للتحقق من القدرة التمييزية للاختبار؛ تم حساب الصدق التمييزى على عينة الدراسة الاستطلاعية البالغ عددها (١٠٠) تلميذ وتلميذة؛ بحساب معامل التمييز لكل مفردة من مفردات

الاختبار بعد ترتيب درجات التلاميذ تنازلياً حسب درجاتهم فى الاختبار، وفصل ٢٧% من درجات أفراد العينة التى تقع فى الجزء الأعلى (الإرباعى الأعلى)، وفصل ٢٧% من درجات أفراد العينة التى تقع فى الجزء الأسفل (الإرباعى الأدنى) ثم استخدام معادلة جونسون لحساب معامل التمييز، وهى: معامل التمييز = $\frac{\text{س}-\text{ص}}{\text{ن}}$ حيث أن: (س–ص) هى الفرق بين عدد الذين أجابوا على المفردة إجابة صحيحة فى كل من المجموعتين العليا والدنيا، (ن) هى عدد أفراد أى من المجموعتين (صلاح علام، ٢٠٠٠: ٢٨٤ – ٢٨٧)، (أمين محمد، رجاء أبوعلام، ٢٠٠٩: ٣١٧).

ويعتبر معامل التمييز للمفردة مقبول إذا زاد عن (٠,٢)، ولذلك فإن الاختبار التحصيلي له القدرة على التمييز بين أفراد العينة. وبعد ذلك تم إعادة ترتيب مفردات الاختبار وفقاً لمعاملات الصعوبة لكل مفردة وجدول (٢) يوضح قيم معاملات التمييز لمفردات الاختبار التحصيلى الموضوعى.

جدول (٢) معاملات التمييز لمفردات الاختبار التحصيلى الموضوعي لمجوعتى الطلاب العليا والدنيا

ID	U	D	$\frac{D-U}{27}$	ID	U	D	$\frac{D-U}{27}$
١	٢٧	٤	٠,٨٥	٥١	٢٣	٢	٠,٧٨
٢	٢٤	٤	٠,٧٤	٥٢	٢٦	٤	٠,٨١
٣	٢٦	١٣	٠,٤٨	٥٣	٢٥	٢	٠,٨٥
٤	٢٦	١٢	٠,٥٢	٥٤	٢٢	٣	٠,٧٠
٥	٢٤	٩	٠,٥٦	٥٥	٢١	٦	٠,٥٦
٦	٢٧	١٠	٠,٦٣	٥٦	٢٥	٧	٠,٦٧
٧	٢٧	١٠	٠,٦٣	٥٧	٢٣	٤	٠,٧٠
٨	٢١	٣	٠,٦٧	٥٨	٢٦	٥	٠,٧٨
٩	٢٢	٨	٠,٥٢	٥٩	٢٣	٤	٠,٧٠
١٠	٢٥	٨	٠,٦٣	٦٠	٢٧	٢	٠,٩٣
١١	٢٣	٨	٠,٥٦	٦١	٢٣	٦	٠,٦٣
١٢	٢٢	٤	٠,٦٧	٦٢	٢٠	١	٠,٧٠
١٣	٢٦	١٤	٠,٤٤	٦٣	٢٧	٦	٠,٧٨
١٤	٢٣	٢	٠,٧٨	٦٤	٢٧	٣	٠,٨٩
١٥	٢٦	٢	٠,٨٩	٦٥	٢٦	١١	٠,٥٦
١٦	٢٤	٣	٠,٧٨	٦٦	٢٣	٦	٠,٦٣
١٧	٢٤	٥	٠,٧٠	٦٧	٢٧	٣	٠,٨٩
١٨	٢٧	١٣	٠,٥٢	٦٨	٢٦	١٣	٠,٤٨

ID	U	D	$\frac{D-U}{27}$	ID	U	D	$\frac{D-U}{27}$
١٩	٢٦	٣	٠,٨٥	٦٩	٢٧	٤	٠,٨٥
٢٠	٢٥	٣	٠,٨١	٧٠	٢٤	٥	٠,٧٠
٢١	٢٣	١	٠,٨١	٧١	٢٦	٣	٠,٨٥
٢٢	٢٧	٨	٠,٧٠	٧٢	٢٣	١٥	٠,٣٠
٢٣	٢٣	١٣	٠,٣٧	٧٣	٢٥	١٤	٠,٤١
٢٤	٢٧	١٥	٠,٤٤	٧٤	٢٣	١٤	٠,٣٣
٢٥	٢٥	٧	٠,٦٧	٧٥	٢٦	٣	٠,٨٥
٢٦	٢٧	١٦	٠,٤١	٧٦	٢٢	١	٠,٧٨
٢٧	٢٦	١٤	٠,٤٤	٧٧	٢٧	٥	٠,٨١
٢٨	٢٤	٣	٠,٧٨	٧٨	٢٥	٥	٠,٧٤
٢٩	٢٦	٣	٠,٨٥	٧٩	٢٤	٤	٠,٧٤
٣٠	٢٣	١	٠,٨١	٨٠	٢٢	٤	٠,٦٧
٣١	٢٥	١٦	٠,٣٣	٨١	٢٤	١٣	٠,٤١
٣٢	٢٢	١١	٠,٤١	٨٢	٢٦	٢	٠,٨٩
٣٣	٢٥	١٣	٠,٤٤	٨٣	٢٧	٣	٠,٨٩
٣٤	٢٥	٥	٠,٧٤	٨٤	٢٠	٢	٠,٦٧
٣٥	٢٤	١٠	٠,٥٢	٨٥	٢٥	٢	٠,٨٥
٣٦	٢٠	١٢	٠,٣٠	٨٦	٢٢	٢	٠,٧٤
٣٧	٢٠	٥	٠,٥٦	٨٧	٢٢	٥	٠,٦٣
٣٨	٢٣	١٣	٠,٣٧	٨٨	٢٦	١	٠,٩٣
٣٩	٢٢	٣	٠,٧٠	٨٩	٢٦	٠	٠,٩٦
٤٠	٢٣	٠	٠,٨٥	٩٠	٢٦	٥	٠,٧٨
٤١	٢٤	١٦	٠,٣٠	٩١	٢٠	٢	٠,٦٧
٤٢	٢٥	٥	٠,٧٤	٩٢	٢١	٢	٠,٧٠
٤٣	٢٣	١٣	٠,٣٧	٩٣	٢٣	٤	٠,٧٠
٤٤	٢٥	٢	٠,٨٥	٩٤	٢٦	٤	٠,٨١
٤٥	٢٢	٣	٠,٧٠	٩٥	٢٦	٥	٠,٧٨
٤٦	٢١	٧	٠,٥٢	٩٦	٢٤	١٤	٠,٣٧
٤٧	٢٥	٢	٠,٨٥	٩٧	٢١	٠	٠,٧٨
٤٨	٢٢	٥	٠,٦٣	٩٨	٢٠	١	٠,٧٠
٤٩	٢٤	٢	٠,٨١	٩٩	٢٦	٥	٠,٧٨
٥٠	٢١	٥	٠,٥٩	١٠٠	٢٢	٧	٠,٥٦

حيث أن Q تشير إلى رقم السؤال، D تشير إلى عدد التلاميذ – من مجموعة التلاميذ منخفضي التحصيل– الذين اجابوا اجابه صحيحة على المفردة، U تشير إلى عدد الطلاب–من

مجموعة التلاميذ مرتفعى التحصيل– الذين اجابوا اجابه صحيحة على المفردة .كما يشير $\frac{D-U}{27}$ إلي قيمة معامل التمييز ويوضح جدول(٨) قيم معاملات التمييز والتى تراوحت بين (. .٣٠ للمفردة رقم ٣٦ إلى. .٩٦ للمفردة رقم ٨٩) وذلك يشير إلى مستوى جيد لقدرة مفردات الاختبار علي التمييز، حيث يذكر (على القرنى، ١٩٩٩، ٥٧) أنه أي مفردة (سؤال) ذات معامل تمييز بين ٢٠ إلى ٣٩% تعتبر ذات تمييز مقبول وأي مفردة ذات تمييز أعلى من ٣٩% تعتبر مفردة جيدة التمييز.

٣-الاتســـاق الـــداخلـــى:

تـم حسـاب الاتسـاق الـداخلي للاختباراسـتخدام برنـامج SPSS علـى عينـة الدراسـة الاستطلاعية (١٠٠) تلميذ وتلميذة من خلال حساب قيمة معاملات الارتباط بين درجة كل مفردة والدرجة الكلية للاختبار بعد حذف درجة المفردة من درجة الاختبار ككل وجدول (٣) يوضح ذلك.

جدول (٣) معاملات الارتباط بين درجة المفردة والدرجة الكلية للاختبار التحصيلى

السؤال	معامل الارتباط	السؤال	معامل الارتباط	السؤال	معامل الارتباط	السؤال	معامل الارتباط
١	٧٠١,٠**	٢٦	٤٢٤,٠**	٥١	٥٩٤,٠**	٧٦	٦٨٢,٠**
٢	٥٨٦,٠**	٢٧	٣٤٩,٠**	٥٢	٦٤٩,٠**	٧٧	٦٤٨,٠**
٣	٣٤٣,٠**	٢٨	٦٢١,٠**	٥٣	٦٦٦,٠**	٧٨	٥٩٦,٠**
٤	٤٥٦,٠**	٢٩	٦٦٧,٠**	٥٤	٥٧٠,٠**	٧٩	٦١٤,٠**
٥	٤٧٢,٠**	٣٠	٦٦٦,٠**	٥٥	٤٦٤,٠**	٨٠	٥١٨,٠**
٦	٤٧٧,٠**	٣١	٣٤١,٠**	٥٦	٥٥٣,٠**	٨١	٣٦٢,٠**
٧	٥٢٨,٠**	٣٢	٣٠٣,٠**	٥٧	٥٣٧,٠**	٨٢	٦٧٦,٠**
٨	٤٥٥,٠**	٣٣	٤٠٤,٠**	٥٨	٥٨٠,٠**	٨٣	٦٧٢,٠**
٩	٣٨٠,٠**	٣٤	٥٦٠,٠**	٥٩	٥١٢,٠**	٨٤	٥٤٩,٠**
١٠	٤٠٥,٠**	٣٥	٤٤٤,٠**	٦٠	٧٢١,٠**	٨٥	٧٠٧,٠**
١١	٣٠٤,٠**	٣٦	٢٠٤,٠*	٦١	٥١٦,٠**	٨٦	٥٦٨,٠**
١٢	٥٠٩,٠**	٣٧	٤٥٥,٠**	٦٢	٥٦٣,٠**	٨٧	٥٠٨,٠**
١٣	٤٢٠,٠**	٣٨	٢٩٣,٠**	٦٣	٦٣٦,٠**	٨٨	٧٣٠,٠**
١٤	٥٥١,٠**	٣٩	٥٢٨,٠**	٦٤	٦٩٨,٠**	٨٩	٧٢٦,٠**
١٥	٦٣٢,٠**	٤٠	٧٥٧,٠**	٦٥	٤٦٦,٠**	٩٠	٥٩٨,٠**
١٦	٥٩١,٠**	٤١	٢٩٥,٠**	٦٦	٥٢٧,٠**	٩١	٥٤٨,٠**
١٧	٥١٥,٠**	٤٢	٥٥٧,٠**	٦٧	٧١٧,٠**	٩٢	٥٨٤,٠**
١٨	٣٨٩,٠**	٤٣	٢٩٣,٠**	٦٨	٤٧٢,٠**	٩٣	٥٧٢,٠**
١٩	٦٤٥,٠**	٤٤	٦٧٢,٠**	٦٩	٧٠٣,٠**	٩٤	٦٥٨,٠**
٢٠	٥٨٨,٠**	٤٥	٥٣١,٠**	٧٠	٥٨٥,٠**	٩٥	٦١٤,٠**
٢١	٦٤١,٠**	٤٦	٤٠٤,٠**	٧١	٦٥٢,٠**	٩٦	٣٢٤,٠**
٢٢	٥٨٦,٠**	٤٧	٥٩٦,٠**	٧٢	٢٨٨,٠**	٩٧	٦٠٧,٠**
٢٣	٢٩٧,٠**	٤٨	٤٦٥,٠**	٧٣	٣٣٣,٠**	٩٨	٥٤٧,٠**
٢٤	٣٧٨,٠**	٤٩	٥٩٥,٠**	٧٤	٢٩٤,٠**	٩٩	٦٢٧,٠**
٢٥	٤٩٥,٠**	٥٠	٤٨٥,٠**	٧٥	٦٥٥,٠**	١٠٠	٤٩٢,٠**

يتضح من جدول (٩) أن قيم معاملات الإرتباط جميعها دالة إحصائياًعند مستوى (٠,٠١) مما يؤكد علي الاتساق الداخلى للاختبار التحصيلى الموضوعى.

٤- معاملات السهولة والصعوبة لمفردات الاختبار التحصيلى الموضوعى.

تم تقدير قيم معامل الصعوبة لكل مفردة من مفردات الاختبار التحصيلى الموضوعى من خلال قسمة عدد الذين أجابوا إجابة صحيحة علي المفردة على العدد الكلي للذين اجابوا على المفردة باستخدام برنامج SPSS وجدول (٤) يوضح تلك القيم والتى تراوحت بين. ٢٤. للمفردة رقم (٤٠) إلى. ٨٨. للمفردة رقم (٢٦) وهى قيم مقبولة إحصائياً وتشير لصدق الأسئلة حيث يُعد السؤال (المفردة) مقبولا إذا تراوحت قيمة معامل الصعوبة له بين (٠,١٥ -. ٨٥.)، كون المفردة التي يقل معامل الصعوبة لها عن. ١٥. تكون شديدة السهولة، والمفردة التي يزيد معامل الصعوبة لها عن. ٨٥. تكون شديدة الصعوبة (صبحى أبوجلالة، ١٩٩٩: ٢٢١) (صلاح الدين علام، ٢٠٠٠: ٢٦٩).

جدول(٤) قيم معاملات الصعوبة لأسئلة الأختبار الموضوعي

السؤال	قيم معاملات الصعوبة	السؤال	قيم معاملات الصعوبة	السؤال	قيم معاملات الصعوبة	السؤال	قيم معاملات الصعوبة
١	٠,٧٧	٢٦	٠,٨٨	٥١	٠,٤٢	٧٦	٠,٢٨
٢	٠,٥٦	٢٧	٠,٦٤	٥٢	٠,٤٢	٧٧	٠,٥٠
٣	٠,٧٥	٢٨	٠,٦٩	٥٣	٠,٥٦	٧٨	٠,٤٣
٤	٠,٧٦	٢٩	٠,٧٠	٥٤	٠,٤٥	٧٩	٠,٣٥
٥	٠,٦٨	٣٠	٠,٣٠	٥٥	٠,٤٥	٨٠	٠,٥١
٦	٠,٧١	٣١	٠,٨٤	٥٦	٠,٥٩	٨١	٠,٦٣
٧	٠,٧٦	٣٢	٠,٥٧	٥٧	٠,٥٧	٨٢	٠,٤٦
٨	٠,٥٨	٣٣	٠,٧٠	٥٨	٠,٤٧	٨٣	٠,٥٢
٩	٠,٦١	٣٤	٠,٦٣	٥٩	٠,٥٠	٨٤	٠,٣٢
١٠	٠,٦٥	٣٥	٠,٦٣	٦٠	٠,٥٧	٨٥	٠,٤٦
١١	٠,٦٣	٣٦	٠,٥٩	٦١	٠,٤٤	٨٦	٠,٤٥
١٢	٠,٦٢	٣٧	٠,٥٩	٦٢	٠,٣٥	٨٧	٠,٤٣
١٣	٠,٨٠	٣٨	٠,٥٥	٦٣	٠,٥٥	٨٨	٠,٤٥
١٤	٠,٥٧	٣٩	٠,٤٥	٦٤	٠,٥٩	٨٩	٠,٤٥
١٥	٠,٦٣	٤٠	٠,٢٤	٦٥	٠,٥٢	٩٠	٠,٥١
١٦	٠,٦٥	٤١	٠,٦٩	٦٦	٠,٤٧	٩١	٠,٣٧
١٧	٠,٦١	٤٢	٠,٥٨	٦٧	٠,٤١	٩٢	٠,٣٤
١٨	٠,٥٩	٤٣	٠,٦٠	٦٨	٠,٦٨	٩٣	٠,٣٧
١٩	٠,٦٤	٤٤	٠,٤١	٦٩	٠,٤٦	٩٤	٠,٤٠
٢٠	٠,٦٥	٤٥	٠,٤٩	٧٠	٠,٤٩	٩٥	٠,٤٨
٢١	٠,٤٠	٤٦	٠,٥٣	٧١	٠,٤٦	٩٦	٠,٥١
٢٢	٠,٧٤	٤٧	٠,٥٣	٧٢	٠,٥٨	٩٧	٠,٣٢
٢٣	٠,٦٨	٤٨	٠,٤٥	٧٣	٠,٥٩	٩٨	٠,٣٠
٢٤	٠,٦٨	٤٩	٠,٥٢	٧٤	٠,٦٢	٩٩	٠,٥٤
٢٥	٠,٥٨	٥٠	٠,٤٠	٧٥	٠,٤٨	١٠٠	٠,٤٥

ثبات الاختبار التحصيلى الموضوعي "التعلق على الجدول".

قامت الباحثة بحساب ثبات الاختبار باستخدام برنامج SPSS حيث كانت قيمة معامل الثبات بعدة طرق هى:

١- باستخدام ألفاكرونباك Cronbach`s Alpha هى (٠,٩٧٦) وهى قيمة مرتفعة يمكن الوثوق بها يتضح من الجدول السابق أن قيمة معامل الثبات للاختبار هى قيمة مرتفعة، مما يدل على ثبات الإختبار وإمكانية الوثوق فى نتائجه.

٢- باستخدام طريقة التجزئة النصفية حيث بلغت قيمة معامل الارتباط بين نصفى الاختبار (٠,٧٣٦) وباستخدام معادلة التصحيح لسبيرمانSpearman-Brown بلغت قيمة معامل الثبات (٠,٨٤٨) وهى قيمة مرتفعة يمكن الوثوق فيها.

الاتساق الداخلي (ثبات المفردات)

قامت الباحثة بحساب معامل الارتباط بين (درجة كل مفردة والدرجة الكلية للاختبار) لكل فرد باستخدام برنامج SPSS ويوضح الجدول (٥) قيم معاملات الارتباط[١]

جدول(٥) معاملات الارتباط بين مفردات الاختبار الموضوعى والدرجة الكلية للاختبار (ثبات المفردات)

السؤال	معامل الارتباط	السؤال	معامل الارتباط	السؤال	معامل الارتباط	السؤال	معامل الارتباط
١	٧٠٩,**	٢٦	٤٣٥,**	٥١	٦٠٦,**	٧٦	٦٩١,**
٢	٥٩٨,**	٢٧	٣٦٥,**	٥٢	٦٦٠,**	٧٧	٦٥٩,**
٣	٣٥٧,**	٢٨	٦٣٢,**	٥٣	٦٧٦,**	٧٨	٦٠٨,**
٤	٤٦٩,**	٢٩	٦٧٦,**	٥٤	٥٨٣,**	٧٩	٦٢٦,**
٥	٤٤١,**	٣٠	٦٧٦,**	٥٥	٤٧٨,**	٨٠	٥٣٢,**
٦	٤٩١,**	٣١	٣٥٣,**	٥٦	٥٦٦,**	٨١	٣٧٨,**
٧	٥٤٠,**	٣٢	٣٢٠,**	٥٧	٥٥٠,**	٨٢	٦٨٦,**
٨	٤٧٠,**	٣٣	٤١٩,**	٥٨	٥٩٣,**	٨٣	٦٨٢,**
٩	٣٩٦,**	٣٤	٥٧٣,**	٥٩	٥٢٦,**	٨٤	٥٦١,**
١٠	٤٢١,**	٣٥	٤٥٩,**	٦٠	٧٣٠,**	٨٥	٧١٦,**
١١	٤١٩,**	٣٦	٢٢٢,*	٦١	٥٣٠,**	٨٦	٥٨١,**
١٢	٥٢٣,**	٣٧	٤٧٠,**	٦٢	٥٧٥,**	٨٧	٥٢٢,**
١٣	٤٣٢,**	٣٨	٣١١,**	٦٣	٦٤٨,**	٨٨	٧٣٨,**
١٤	٥٦٤,**	٣٩	٥٤٢,**	٦٤	٧٠٨,**	٨٩	٧٣٥,**
١٥	٦٤٣,**	٤٠	٧٦٤,**	٦٥	٤٨١,**	٩٠	٦١٠,**
١٦	٦٠٣,**	٤١	٣١١,**	٦٦	٥٤٠,**	٩١	٥٦١,**
١٧	٥٢٨,**	٤٢	٥٧٠,**	٦٧	٧٢٦,**	٩٢	٥٩٦,**
١٨	٤٠٥,**	٤٣	٣١٠,**	٦٨	٤٨٦,**	٩٣	٥٨٥,**
١٩	٦٥٦,**	٤٤	٦٨٣,**	٦٩	٧١٣,**	٩٤	٦٦٨,**
٢٠	٥٩٩,**	٤٥	٥٤٥,**	٧٠	٥٩٨,**	٩٥	٦٢٦,**
٢١	٦٥٢,**	٤٦	٤٢٠,**	٧١	٦٦٣,**	٩٦	٣٤١,**
٢٢	٥٩٧,**	٤٧	٦٠٨,**	٧٢	٣٠٥,**	٩٧	٦١٨,**
٢٣	٣١٣,**	٤٨	٤٨٠,**	٧٣	٣٥٠,**	٩٨	٥٥٩,**
٢٤	٣٩٤,**	٤٩	٦٠٨,**	٧٤	٣١١,**	٩٩	٦٣٩,**
٢٥	٥٠٩,**	٥٠	٤٩٩,**	٧٥	٦٦٥,**	١٠٠	٥٠٦,**

(١) تعبر عن قيم جميعها دالة أحصائيا عن وجدود اتساق داخلى بين درجات الطلاب على كل مفردة والمجموع الكلى لدرجات الطلاب على نفس المفردات.

يتضح من جدول (١١) أن قيم معاملات الأرتباط دالة احصائياً[١] عند مستوى (٠,٠١)**، وعند مستوى (٠,٠٥)*

حساب زمن الإختبار:

تم تحديد الزمن اللازم لتطبيق الاختبار التحصيلى، عن طريق حساب المنوال، فتم حساب منوال الأزمنة التى استغرقتها كل تلميذة من تلميذات العينة الاستطلاعية فى الإجابة عن مفردات الاختبار، حيث بلغ عدد من أنهى الاختبار فى زمن ٦٠ دقيقة هو(٤٠) ممتحن، (٦٠) ممتحن فى زمن ٩٠ دقيقة، بناءً على ذلك فإن الزمن اللازم للإجابة عن مفردات الاختبار هو(٩٠) دقيقة.

إجراءات الدراسة:

سارت إجراءات الدراسة الحالية على النحوالتالى:

١- تم تطبيق أدوات الدراسة على العينة الاساسية البالغ عددها (٤٧٦) تلميذ وتلميذة بالصف الثانى الأعدادى.

٢- تم تصحيح أداة الدراسة وإعداد ملف البيانات بهدف معالجتها إحصائياً.

الأساليب الاحصائية:

تم استخدام البرنامج الإحصائى "WINSTEPS"بهدف استبعاد الأفراد والمفردات غير الملائمة إحصائياً لأسس القياس الموضوعى وأيضاً تدريج المفردات، وتقدير قدرات الأفراد المقابلة لكل درجة خام احتمالية وإيجاد قيمة دالة المعلومات لكل اختبار.

نتائج الدراسة ومناقشتها:

يُعرض فيما يلى نتائج الدراسة ومناقشتها:

نتائج فرض الدراسة: ينص الفرض على أنه "تختلف تقديرات معلم صعوبة المفردة ومعلم قدرة الفرد باختلاف طول الاختبار الموضوعى المستخدم (٢٨ – ٥٣ – ١٠٠) مفردة"، وللتحقق من صحة الفرض الرابع: استخدمت الباحثة نموذج راش لتحليل البيانات من خلال البرنامج الإحصائى (Winsteps) لتدريج مفردات الاختبار الموضوعي **باختلاف طول الاختبار الموضوعى المستخدم (٢٨ – ٥٣ – ١٠٠) مفردة** ومنه يتضح ما يلى:

(١) قيمة معامل الارتباط الجدولية بدرجات حرية ٩٨ عند مستوى دلالة ٠,٠١ =٠,٢٥٤، وعند مستوى دلالة ٠,٠٥ = ٠,١٩٥

أولاً: بالنسبة لتحليل الاختبار الموضوعى باستخدام اختبار طوله (٢٨ مفردة).

١- يمتد مدى الصعوبة الذى تغطيه مفردات الاختبار من (-٤,١٥: ٢,٢) لوجيت أى (٢٩,٢٥: ٦١) منف.

٢- تمتد الأخطاء المعيارية لتقديرات الصعوبة بين (٠,١١: .٣٧٠) لوجيت، وتعتبر قيم الأخطاء المعيارية – هذه – صغيرة نسبياً مما يدل على دقة وثبات تقديرات صعوبات المفردات.

٣- يتراوح تقديرات القدرة المقابلة لكل درجة كلية محتملة خام على مقياس الاسئلة الكلى بين (-٥,٥٧: ٤,٧٧) لوجيت، أي من (٢٢,١٥: ٧٣,٨٥) منف، وتعتبر هذه الحدود أقصى حدود للقدرة يمكن تقديرها.

٤- يمتد مدى قدرة الافراد التى تغطيه عينة الاختبار من (-٥,٥٧: ٤,٧٧) لوجيت أى (٢٢,١٥: ٧٣,٨٥) منف.

٥- مقدار المعلومات الذى يعطيها الاختبار الموضوعى القصير عند مستوى قدرة (θ) يساوى (٣,٥) وهى أكبر كمية معلومات يعطيها هذا الاختبار والذى يتحدد عند مستوى قدرة فى الفتري من (-٢: ٣) لوجيت.

- ويمكن حساب قيمة الخطأ المعياري لهذا الاختبار باستخدام المعادلة التالية: $S.E(\theta) = \frac{1}{\sqrt{I(\theta)}} = \frac{1}{1.9} = 0.53$ وتعد تلك القيمة للخطأ المعيارى أقل مايمكن عند نفس الفترة من (-٢:٣).

ثانياً: بالنسبة لتحليل الاختبار الموضوعى باستخدام اختبار طوله (٥٣ مفردة).

١- يمتد مدى الصعوبة الذى تغطيه مفردات الاختبار من (-٢,٣٦: ٥,٤٥) لوجيت أى (٣٨,٢: ٧٧,٢٥) منف.

٢- تمتد الأخطاء المعيارية لتقديرات الصعوبة بين (٠,١١: .٣٦٠) لوجيت، وتعتبر قيم الأخطاء المعيارية – هذه – صغيرة نسبياً مما يدل على دقة وثبات تقديرات صعوبات المفردات.

٣- يتراوح تقديرات القدرة المقابلة لكل درجة كلية محتملة خام على مقياس الاسئلة الكلى بين (-٥,٥١: ٦,٦٤) لوجيت، أي من (٢٢,٤٥: ٨٣,٢) منف، وتعتبر هذه الحدود أقصى حدود للقدرة يمكن تقديرها.

٤- يمتد مدى قدرة الافراد التى تغطيه عينة الاختبار من (-٣,٥١: ٤,٩١) لوجيت أى (٣٢,٤٥: ٧٤,٥٥) منف.

٥- مقدار المعلومات الذى يعطيها الاختبار الموضوعى متوسط الطول عند مستوى قدرة (θ) يساوى (٨) وهى أكبر كمية معلومات يعطيها هذا الاختبار والذى يتحدد عند مستوى قدرة فى الفتري من (-١: ٢) لوجيت.

- ويمكن حساب قيمة الخطأ المعياري لهذا الاختبار باستخدام المعادلة التالية: $S.E(\theta)=\frac{1}{\sqrt{I(\theta)}}=\frac{1}{2.8}=0.35$ وتعد تلك القيمة للخطأ المعيارى أقل مايمكن عند نفس الفترة من (-١: ٢).

ثالثاً: بالنسبة لتحليل الاختبار الموضوعى باستخدام اختبار طوله (١٠٠ مفردة).

١- يمتد مدى الصعوبة الذى تغطيه مفردات الاختبار من (-٣,٩: ٥,٧١) لوجيت أى (٣٠,٤٥: ٧٨,٧٥) منف.

٢- تمتد الأخطاء المعيارية لتقديرات الصعوبة بين (٠,١٢: .٣٩.) لوجيت، وتعتبر قيم الأخطاء المعيارية – هذه – صغيرة نسبياً مما يدل على دقة وثبات تقديرات صعوبات المفردات.

٣- يتراوح تقديرات القدرة المقابلة لكل درجة كلية محتملة خام على مقياس الاسئلة الكلى بين (-٦,٢٧: ٧,٠١) لوجيت، أي من (١٨,٦٥: ٨٥,٠٥) منف، وتعتبر هذه الحدود أقصى حدود للقدرة يمكن تقديرها باستخدام الاختبار الموضوعى.

٤- يمتد مدى قدرة الافراد التى تغطيه عينة الاختبار من (-٣,٤: ٥,٤١) لوجيت أى (٣٣: ٧٧,٠٥) منف.

٥- مقدار المعلومات الذى يعطيها الاختبار الموضوعى عند مستوى قدرة (θ) يساوى (١٤,٥) وهى أكبر كمية معلومات يعطيها هذا الاختبار والذى يتحدد عند مستوى قدرة فى الفتري من (-١: ١) لوجيت.

- ويمكن حساب قيمة الخطأ المعياري لهذا الاختبار باستخدام المعادلة التالية: $S.E(\theta)=\frac{1}{\sqrt{I(\theta)}}=\frac{1}{3.8}=0.26$ وتعد تلك القيمة للخطأ المعيارى أقل مايمكن عند نفس الفترة من (-١:١).

منـاقشـة نتـائـج فـرض الــدراســة

جدول (٦) جدول يوضح ملخص نتائج تدريج عينة مكونة من ٤٧٦ تلميذ وتلميذة باستخدام أطوال مختلفة من الاختبار الموضوعى (٢٨ – ٥٣ – ١٠٠) مفردة

طول الاختبار الموضوعى المستخدم فى التحليل	مدى صعوبة المفردات المكونه للاختبار	مدي الاخطاء المعيارية لتقديرات صعوبة المفردات	مدى قدرة الافراد التى يقيسها الاختبار	مدى تقديرات القدرة المقابلة لكل درجة كلية	دالة معلومات الاختبار	الخطأ المعيارى للاختبار
اختبار قصير (٢٨ مفردة)	٦١:٢٩,٢٥	٣٧٠.:٠,١١	٧٣,٨٥:٢٢,١٥	٧٣,٨٥:٢٢,١٥	٣,٥	٠,٥٣
اختبار متوسط (٥٤ مفردة)	٧٧,٢٥:.٣٨,٢	٣٦٠.:٠,١١	٧٤,٥٥:٣٢,٤٥	٨٣,٢:٢٢,٤٥	٨	٠,٣٥
اختبار طويل (١٠٠ مفردة)	٧٨,٧٥:٣٠,٤٥	٣٩٠.:٠,١٢	٧٧,٠٥:٣٣	٨٥,٠٥:١٨,٦٥	١٤,٥	٠,٢٦

بالنظر إلى جدول (٦) يتضح اختلاف تقديرات معلم صعوبة المفردة ومعلم قدرة الفرد ودالة معلومات الاختبار باختلاف طول الاختبار الموضوعى المستخدم (٢٨ – ٥٣ – ١٠٠) مفردة، حيث شملت مفردات الاختبار الموضوعي باستخدام اختبار **متوسط الطول** مدى أكبر لصعوبة المفردات المكونه للاختبار وأقل قيم للاخطاء المعيارية لصعوبة المفردات المكونه للاختبار، أظهرت مفردات الاختبار الموضوعي باستخدام **اختبار قصير** أعلى مدى لقدرة الافراد التى يقيسها الاختبار، بينما أظهرت مفردات الاختبار الموضوعي باستخدام اختبار **طويل** مدى أعلى لتقديرات القدرة المقابلة لكل درجة كلية وأعلى قيمة لدالة معلومات الاختبار وأقل قيمة لمقدار الخطأ المعيارى للاختبار.

علاقة نتيجة الفرض بنتائج الدراسات السابقة

تختلف دراسة Bastari (2000) ودراسة **الشيماء عبد الحميد محمد (٢٠٠٧)** مع نتيجة الفرض الرابع فى أنه بزيادة طول الاختبار يزداد دقة بارامترات المفردة المكونه للاختبار.

وتتفق مع دراسة **(Sahin, A. & Weiss, D., 2015)** والتى تشير إلى أن زيادة طول بنك المفردة يحدث تأثير إيجابي على تقديرات قدرة الأفراد، كما أن المزيد من المفردات في بنك المفردة ينتج عنه مزيد من المعلومات.

وتتفق نتائج دراسة **الشيماء عبد الحميد محمد (٢٠٠٧)** مع الفرض الرابع فى أن أفضل مطابقة للمفردات مع نموذج راش تتحقق عند الاختبار متوسط الطول.

التوصيات المقترحة: يمكن الخروج بالتوصيات التالية بناء على نتائج الدراسة الحالية:

١- اسفرت الدراسة عن نتائج أستخدام أطوال عدة للاختبارات يزود مستخدمي نظرية الاستجابة للمفردة بتوضيح تفصيلي حول أطوال للاختبار من أجل أستفادة مثلى من عملية القياس.

٢- توصى الباحثة فى ضوء نتائج البحث الحالى باستحدام اختبارات طويلة مما يحقق خصائص سيكومترية أدق.

المـــراجـــــع

أولاً: المـراجـع العـربيـة:

١- أحمد إسماعيل الكحلاوى، هند عبد المجيد الحمورى (١٩٩٨). أستراتجيات بناء بنك الأسئلة فى الرياضيات. **مجلة كلية التربية، جامعة الأمارات**،١٣ (١٥) ،٤٧: ٩٤.

٢- اعتدال غازى عبدالله (٢٠٠٠). القياس محكى المرجع لأحد مقررات علم النفس التعليمى، **رسالة ماجستير** غير منشورة، كلية التربية للبنات، جامعة عين شمس.

٣- الشيماء عبد الحميد محمد (٢٠٠٧). أثر كل من حجم العينة وطول الاختبار على دقة نتاج الاختبارات التحصيلية مرجعية المحك فى الرياضيات باستخدام نموذج راش، **رسالة ماجستير** غير منشورة٠ كلية التربية، جامعة الزقازيق.

٤- أمين على محمد سليمان، رجاء محمود أبوعلام (٢٠٠٩). **القياس والتقويم فى العلوم الإنسانية أسسة وأدواته وتطبيقاته**٠ القاهرة: دار الكتاب الحديث.

٥- أمينة محمد كاظم, أنور محمد الشرقاوى, سليمان الخضرى الشيخ ونادية عيد السلام (١٩٩٦).**اتجاهات معاصرة فى القياس والتقويم النفسى والتربوى**. القاهرة: مكتبة الانجلو المصرية.

٦- خالد بن حسن شيبان التميمى (١٩٩٩). أثر كل من نوع المحكم وطول الاختبار على تحديد درجة القطع لاختبار محكى المرجع يقيس الكفايات الرياضية فى العمليات الحسابية على الأعداد بالصف السادس الابتدائى بمدينة جدة، بحث مكمل لمتطلبات نيل درجة الماجستير. كلية التربية. جامعة أم القرى.

٧- رضا سمير عوض (٢٠٠٤). دراسة أثر عدد المفردات وموقع المفردة من الاختبار وحجم العينة وعدد البدائل على قيمة معامل ثبات القياس، **رسالة ماجستير** غير منشورة، معهد الدراسات التربوية، جامعة القاهرة.

٨- صلاح أحمد مراد، أمين على سليمان (٢٠٠٥). **الاختبارات والمقاييس فى العلوم النفسية والتربوية خطوات إعدادها وخصائصها**. ط٢. القاهرة: دار الكتاب الحديث.

٩- صلاح الدين محمود علام (١٩٨٧). دراسة موازنه ناقدة لنماذج السمات الكامنه والنماذج الكلاسيكية فى القياس النفسي والتربوى. **المجلة العربية للعلوم الأنسانية**، ٧ (٢٧)، ١٨-٤٣.

١٠- صلاح الدين محمود علام (٢٠٠٠). **القياس والتقويم النفسى والتربوى " أساسياته وتطبيقاته وتوجهاته المعاصرة"**. القاهرة: دار الفكر العربى.

١١- صلاح الدين محمود علام(٢٠٠٥). **نماذج الاستجابة للمفردة الاختبارية احادية البعد ومتعددة الابعاد وتطبيقاتها فى القياس النفسى والتربوى** . القاهرة: دار الفكرالعربى.

١٢- صلاح الدين محمود علام (٢٠٠٦). **الاختبارات والمقاييس النفسية والتربوية**. عمان: دار الفكر للنشر.

١٣- صلاح الدين علام (٢٠٠٦). **القياس والتقويم النفسى والتربوى، أساسياته وتطبيقاته وتوجهاته المعاصرة**. القاهرة: دار الفكر العربى.

١٤- صلاح الدين محمود علام (٢٠٠٧). **القياس والتقويم التربوى فى العملية التدريسية**.عمان: دار المسيرة.

١٥- طلال فرج يوسف كيلانى(٢٠١٢).الاستخدام الامثل لوسائل القياس والتقويم ودورها فى ضمان جودة مخرجات التعليم الجامعى .**المجلة العربية لضمان جودة التعليم الجامعى**, ٥ (٩), ٢٢.٤٥.

١٦- محمد إبراهيم محمد محمد (٢٠١٠). استخدام نموذج التقدير الجزئى لتحليل مستويات بنية ناتج التعلم والاختبارات العنقودية ذات الفقرات المرتبة هرمياً لدى عينة من طلاب كلية التربية بالمنيا. **رسالة دكتوراة غير منشورة,** كلية التربية, جامعة المنيا.

١٧- محمد منصور الشافعى (١٩٩٦). أثر طرق معادلة درجات الاختبار وضوابط اختيار العينة على تدريج بنود بنك الاسئلة. **رسالة دكتوراة غير منشورة،** كلية التربية، جامعة المنصورة.

١٨- نضال كمال الشريفين (٢٠٠٦)، الخصائص السيكومترية لاختبار محكى المرجع فى القياس والتقويم التربوى وفق النظرية الحديثة فى القياس التربوى والنفسى . **مجلة العلوم التربوية والنفسية، ٧ (٤)، ٨١-١٠٧٠.**

١٩- هشام فتحى محمد جاد الرب (١٩٩٩). تطوير اختبار للذكاء باستخدام نماذج السمات الكامنه وأثر ذلك على قدرة الاختبار على التنبؤ بالتحصيل الدراسى، **رسالة ماجيتير**، جامعة المنصورة.

٢٠- يحى علوان(٢٠٠٧).التقويم والقياس التربوى ودورة فى انجاح العملية التربوية. **مجلة العلوم الانسانية**, جامعة محمد خضير بسكرة, ١١، ٩.٣١.

ثانياً: المـــراجـــع الإجنبيـــة:

1- Abd.razak, n.&khairani, a. and meethien, l.(2012). Examining quality of mathematics test items using rasch model:preminarily analysis, available on line at **http://www.science direct.com.**

2- Akour, m. and Al-omari, h.(2013). Empricical investigation of the stability of irt item .**internationalon line journal of education sciences**, vol.5, no.2, P.p .291-301 .

3- Baker, F.B. (2001).**The Basics Of Item Response Theory**. United States Of America.Port smouth, NH:Heinemann Educational Books.

4- Bastari, B. (2000). Linking multiple-choice and constructed-response items to a common proficiency scale.**Submitted to the graduate school of the university of Massachusetts Amherst in partial fulfillment of the requirements for the degree of doctor of education.**

5- Cunningham, G. (1986). **Educational and Psychological Measurement**. Macmillan Publishing Company New York, Collier Macmillan Publishers London.

6- Gleason, J.(2008).An Evaluation of Mathematics Competitions Using Item Response Theory . **Notices Of The Ams** , Vol.55, no.1, Pp.8-15.

7- Hidalgo, M. & Martinez, M. & Benito , J. and Guilera, G. (2016) . Acomparison of discriminant logistic regression and item response theory likelihood-Ratio tests for differential item functioning (IRTLRDIF) in polytomouse short tests, **Psicothema.** Vol.28, No.1, Pp.83-88.

8- Jiang, S.&Wang, C. and Wess, D.(2016).**Frontiers in psychology.**Vol.7, February.

9- Kesamang, M.(2017).The comparison of item and person item response theory (IRT) parameter estimations for the anchor-items and common-persons designs, **Herald Journal of educaton and general studies** . Vol. 3, No. 1, Pp.1-12, March.

10- Oji, E. & Charles, I. (2016) . Transformation from classical test theory (CTT) to item response, **URDO-Journal of education reseach .** Vol.1, No.4, Pp.104:116.

11- Sahin, A. and Weiss, D. (2015).Effects of calibration sample size and item bank siz on ability estimation in computerized adaptive testing. **Educational science theory &Practice,** December 2015, 15(6), 1585-1595.

12- Yousfi, S. (2012). Principles and Procedures of Considering Item Sequence effects in the development of calibrated item pools: conceptual analysis and empirical Mutation, **psychological test and assessment Modeling**, vol. 54, No. 4, Pp. 60- 396