المفاضلة بطرق تصحيح اختبار الصواب والخطأ المتعدد في الخصائص القياسية لاختبارات تحصيلية

أ. م. د. هند صبيح رحيم
قسم العلوم التربوية والنفسية- كلية التربية
ابن رشد للعلوم الانسانية- جامعة بغداد
hind.sabeeh@ircoedu.uobaghdad.edu.iq

المستخلص :
هدفت الدراسة المفاضلة بطرق تصحيح اختبار الصواب والخطأ المتعدد في الخصائص القياسية للاختبارات تحصيلية وغرض تحقيق الهدف تم بناء اختبار تحصيلي في مادة منهج البحث العلمي مكون من (25) فقرة ذات البدائل الاربعة لاختبار الصواب والخطأ المتعدد وتم تطبيق الاختبار على عينة (400) طالب وطالبة من الصف الثالث من طلبة كليات التربية في بغداد، وتبين ان الاختبار وفقراته يتمتع بخصائص قياسية جيدة ،واستعملت الباحثة اربع طرق من طرق تصحيح اختبار الصواب والخطأ المتعدد المختلفة ،والتعرف على دلالة الفروق لطرق الاربعة لتصحيح الاختبار توصلت الباحثة الى ،عدم وجود فروق ذات دلالة احصائية في معاملات صعوبة الفقرات بين الطرق الاربعة، اظهرت النتائج وجود فروق في القوة التمييزية بين الطرق لصالح الطريقة الرابعة ، واظهرت النتائج وجود فروق في معاملات ارتباط الفقرات بالدرجة الكلية للاختبار بين الطرق لصالح الطريقة الرابعة ، اظهرت النتائج وجود فروق في الثبات بين الطرق لصالح الطريقة الرابعة. وتعد طريقة التصحيح الرابعة هي افضل الطرق التصحيح وذلك حيث تمنح الطالب الدرجة الاعلى فضلا عن مراعاتها المعرفة الجزئية للطالب وبعد ذلك تأتي الطريقة الثالثة والثانية ثم الاولى قدمت الباحثة عدد من التوصيات والمقترحات .

الكلمات المفتاحية : الاختبار، الصواب والخطأ المتعدد، الخصائص القياسية

Comparison of methods of correcting multiple true and false tests in the standard characteristics of achievement tests

Asst. Prof. Dr. Hind Sabeeh Rahim
Department of Educational and Psychological Sciences
Ibn Rushd College of Education for Human Sciences University of Baghdad
hind.sabeeh@ircoedu.uobaghdad.edu.lg

The study aimed to differentiate methods for correcting the multiple true and false test in the standard characteristics of achievement tests. For the purpose of achieving the goal, an achievement test was constructed in the Scientific Research Methodology subject, consisting of (25) items with four alternatives to the multiple true and false test. The test was applied to a sample of (400) male and female students from The third year of students in colleges of education in Baghdad, and it turned out that the test and its items have good standard characteristics. The researcher used four different methods of correcting the multiple true-false test, and to identify the significance of the differences for the four methods of correcting the test. The researcher concluded that there are no statistically significant differences in Parameters of difficulty of items between the four methods. The results showed that there were differences in discriminatory power between the methods in favor of the fourth method. The results showed that there were differences in the correlation coefficients of the items with the total score of the test between the methods in favor of the fourth method. The results showed that there were differences in reliability between the methods in favor of the fourth method. The fourth method of correction is the

best method of correction, as it gives the student the highest grade in addition to taking into account the student's partial knowledge. After that comes the third method, the second, and then the first. The researcher presented a number of recommendations and proposals.

Keywords: testing, multiple true and false, standard properties

مشكلة البحث The problem of the Research :

هناك العديد من الاختبارات التحصيلية التي تعد اكثر ادواة القياس استعمالا في قياس محصلة التعليم ،"يشير مفهوم التحصيل الى درجة ا و مستوى النجاح ا لذي يحرزة الطالب وفي مجال دراسي عام او متخصص ، فهو" يمثل اكتساب المعارف والمهارات والقدرة على استخدامها في مواقف حالية او مستقبليه ، ويعد التحصيل الناتج النهائي للتعليم ،وعلى الرغم من دور الاختبارات في العملية التربوية ،الا ان هناك وجهات نظر مختلفة حول طبيعتها واشكالها وطرق اجرائها ،وكذلك صلتها با الاهداف التربوية وطرق تدريسها "(علام ،2014: 122) لذلك احرزت الاختبارات التحصيلية في المدارس والجامعات الدرجة الاولى لذلك تتولى المؤسسات التربوية عناية خاصة بالاختبارات التحصيلية . وجدت الباحثة من خلال خبرتها في التدريس ان هناك مشاكل وبعض الصعوبات في الاختبارات التحصيلية التقليدية لذلك وجدت الباحثة هذه المشكلة لغرض معالجة ومنها طرق تصحيح اختبار الصواب والخطأ المتعدد الذي يعد نوع متقدم من انواع الاختبارات التحصيلية الموضوعية التي يحتاج تحديد أي الطرق افضل وذلك لوجود عدد من طرق الصحيح في تحقيق الخصائص القياسية لكل من فقرات الاختبار والاختبار ككل ، هناك عدد من الدراسات العربية تناولت طرق التصحيح اختبار الصواب والخطأ المتعدد منها دراسة(سوالمة 1994)ودراسة (الصمادي،2012) حيث الدراسات التي تناولت طرق تصحيح اختبار الصواب والخطاء المتعدد .ومن خلال النتائج التي توصلت اليها الدراسات السابقة ووجود طرق متعددة لتصحيح الاختبار الصواب والخطأ المتعدد كان على الباحثة من خلال هذه الدراسة الكشف عن هذه الطرق وايهما تتمتع بخصائص قياسية افضل للفقرات والاختبار ،حسب علم الباحثة تعد هذه الدراسة الاولى في العراق في المفاضلة بطرق تصحيح اختبار الصواب والخطأ المتعدد .

أهمية البحث The significance of the Research :

قد اسهم التطور المتسارع في منهجيات وتقنيات علم القياس في تحديث وطرق واساليب بناء مختلفة الى انواع وادوات قياس ،بما يؤدي الى زيادة دقة وموضوعية قياس وتقويم السمات الانسانية المتعددة (علام ،2006: 11). ومن المسلم به ان زيادة الخصائص السيكو مترية للاختبار تؤدي الى ارتفاع درجة الاطمئنان للقرارات التي تبنى على نتائج القياس ،"يهتم القائمون ببناء الاختبارات والمقاييس بكتابة او انتقاء فقرات عالية الجودة لقياس سمات الانسانية قياسا دقيقا ،لذلك يراعون كثيرا من الشروط في تكوين هذه الفقرات وصياغتها والتحقق بالأساليب المنطقية واحكام الخبراء من صدق المحتوى كل فقرة على حدة وكذلك محتوى الاختبار ككل . غير انه مهما بلغت هذه الاساليب والاحكام فأنها لا تغني عن التجريب الميداني للاختبار وتحليل درجات فقراته باستخدام الاساليب الاحصائية وتحديد العلاقة بين ما تقيسه الفقرات وبين استجابة الافراد لها" (علام ،2011: 267).ذكر (النبهان 2013)"ان اختبار الصواب والخطأ المتعدد ما هو الاحالة خاصة من فقرات اختيار من متعدد، واكتسبت الاختبارات الموضوعية شهرة ونتشار كبيرين لما تتمتع به من الموثوقية والبعد عن التحيز في اعطاء الدرجات ،ورغم ذلك تواجه النقد في بعض الحالات ، كما تقابل بالتأييد والاستحسان في حالات اخرى "(عمر ،واخرون ،2010: 379).تبرز اهمية هذه الدراسة من خلال توفر اختبار تحصيلي يتمتع بخصائص قياسية لمادة منهج البحث العلمي ، فضلا عن توفر الادلة حول ماهية الاسلوب الانسب للحصول على اختبار يتمتع بخصائص قياسية جيدة (صدق وثبات)وعلى فقرات جيدة من حيث الصعوبة والتمييز .ان اهم ما يميز هذا الاختبار هو قدرته على تحقيق التوازن ما بين المحتوى الدراسي والاهداف التعليمية التي تسعى المؤسسة التعليمية الى الوصول اليها ،كذلك تمكن المستفيدين من هذا الاختبار من تحديد افضل

وانسب طرق تصحيح اختبار الصواب والخطأ المتعدد مما يشجع على استخدام هذا النوع من الاختبارات لما له من خصائص جيدة يتمتع به الاختبار وفقراته بدلا من الاختبارات الاخرى التقليدية في التحصيل .

هدف البحث: Aim of the Research:

يهدف البحث الحالي الى:

1- بناء اختبار تحصيلي الصواب والخطأ المتعدد لمادة منهج البحث العلمي للصف الثالث لطلبة كليات التربية في بغداد

2- التعرف على دلالة الفروق في الخصائص القياسية للاختبار التحصيلي في مادة منهج البحث العلمي باختلاف طرق تصحيح اختبار الصواب والخطأ المتعدد (ط1، ط2، ط3، ط4) ويشتق من هذا الهدف الفرضيات التالية :

- **الفرضية الاولى:** لا توجد فروق ذات دلالة احصائية بين معاملات صعوبة الفقرات المحسوبة للطرائق الاربعة في تصحيح اختبار الصواب والخطأ المتعدد (ط1 – ط2 – ط3 – ط4).
- **الفرضية الثانية:** لا توجد فروق ذات دلالة احصائية بين معاملات تمييز الفقرات المحسوبة للطرائق الاربعة في تصحيح اختبار الصواب والخطأ المتعدد (ط1 – ط2 – ط3 – ط4).
- **الفرضية الثالثة :** لا توجد فروق ذات دلالة احصائية بين معاملات ارتباط الفقرات بالدرجة الكلية للاختبار المحسوبة للطرائق الاربعة في تصحيح اختبار الصواب والخطأ المتعدد(ط1 – ط2 – ط3 – ط4) .
- **الفرضية الرابعة :** لا توجد فروق ذات دلالة احصائية بين معاملات الثبات المحسوبة للطرائق الاربعة في تصحيح اختبار الصواب والخطأ المتعدد(ط1 – ط2 – ط3 – ط4)

حدود البحث: Limitation of the Research:

1- يتحدد البحث الحالي بطلبة الصف الثالث لكليات التربية في محافظة بغداد و لكلا الجنسين (ذكور– اناث) .

2. مادة منهج البحث العلمي للصف الثالث لطلبة كليات التربية .

تحديد المصطلحات : Definition of terms :

1- الاختبار Test:- عرفة كل من:

- **كرونباخ(Cronback1970):-**

" ذلك الاختبار الذي حددت إجراءاته وإداراته ورتبت وحداته وتأكد ما ثباته بحيث يصلح للتطبيق في زمان ومكان معينين" (Cronback, 1970: 22).

- **عودة (2000) :**

"طريقة لتقدير درجة امتلاك الفرد لسمة معينة من خلال إجابات الفرد مع على عينة من المثيرات التي تمثل السمة" (عودة، 2000: 36).

*** محاسنة(2013)**

" اجراء منظم لقياس عينة من السلوكيات من خلال عينة من المثيرات ،اي انه اداة للحصول على معلومات حول سلوك الفرد ونقول عينة لانه لا يمكن الحصول على جميع سلوكيات الفرد وبالتالي لا يمكن تحديدها وبالتالي اخضاعها للاختبار وتصمم الاختبارات عادة لا اغراض مختلفة" (محاسنة، 2013 : 51) .

– **لصواب والخطأ المتعدد:** عرفة كل من :

- **(Ebel & Frisbie،1991):**

" تعد شكلا من أشكال فقرات الاختيار من متعدد وتتكون من متن واحد وعدة بدائل بعضها أو كلها صحيحة أو خطأ، ويطلب فيها من المستجيب الاستجابة لكل بديل من البدائل على حدة بوضع علامة (√) أو علامة (×) "(Ebel & Frisbie،1991:3).

•ابو جراد(2015):

" فقرة عنقودية تتكون من متن ومجموعة من البدائل التي تكملها بصورة مستقلة وكل بديل عبارة فقرة صواب وخطأ"(ابو جراد، 2015 : 3).

التعريف الاجرائي :هي الدرجة التي يحصل عليها المستجيب على فقرات الصواب والخطأ المتعدد وتعطي درجة وفقاً لكل طريقه من طرق تصحيح الاختبار الاربعة المعدة لهذه الدراسة .

الخصائص القياسية :عرفها كل من :

مراد وسلمان (2005):

" مجموعة من الخصائص الضرورية والمتعلقة بالصدق والثبات والمعايير التي يتحتم احتسابها بعد التجريب الاختبار على عينة ممثلة للمجتمع وتعتمد جودة الاختبار على مدى توفر بيانات مناسبة لهذه الخصائص "(مراد وسلمان، 2005: 350).

عزيز (2015) :

" مجموعة من الشروط اللازمة لكي تكون اداة القياس صالحة لقياس السمة المستهدفة، وهي بمثابة دلائل او مؤشرات احصائية عن مدى جودة الاداة وما تتضمن من فقرات ،وعن مدى الوثوق والاطمئنان لنتائجها"(عزيز، 2015 :22).

الفصل الثاني: اطار نظري ودراسات سابقة

اولا : الاطار النظري

الاختبارات التحصيلية :

تلعب الاختبارات التحصيلية، وما زالت الوسيلة الأساسية في تقويم العملية التربوية للتعرف على ما حققه الطالب من اهداف تعليمية، والاختبارات كأداة من أدوات التقويم التربوي تعد من اقدم ادوات التقويم المستخدمة وأكثرها شيوعا وقبولا عند أولياء الامور، ومع ان الاتجاهات في التقويم التربوي تشجع على استخدام اساليب اخرى غير الاختبارات التحصيلية مثل ملفات الانجاز ألا انه لا يزال للاختبارات التحصيلية دور مهم في عملية التقويم التربوي و دورا اساسيا في حياة الطالب في جميع مراحل التعليم، وبناء على نتائجها يتحدد مستقبل الطالب، وتتخذ بحقه الكثير من القرارات "(ابو فوده، وبني يونس، 2012: 24).

الأسس العامة للاختبارات التحصيلية :

1. "تقيس المعطيات التعليمية المحددة التي تتناغم مع الأهداف التعليمية، وهذه المعطيات تتمثل بمعرفة الحقائق المحددة، ومعرفة المصطلحات، وفهم المفاهيم والأُسس، والقدرة على تطبيق الحقائق والأُسس وبعض مهارات التفكير الأخرى.
2. تقيس عينة ممثلة للمعطيات التعليمية وللمادة التعليمية .
3. تتضمن أنواعاً من الفقرات الأكثر ملاءمة لقياس المعطيات التعليمية المتطلبة وتقويمها.
4. تصمم لتلائم الاستخدامات المعنية بنتائجها.
5. ثابتة وقابلة للتفسير.
6. صادقة في تحقيق التعلم "(البغدادي، 1984 : 112-119).

استخدامات الاختبارات والمقاييس التربوية والنفسية :

تستخدم الاختبارات التربوية والنفسية في المعاونة في اتخاذ القرارات المختلفة ، وفيما يأتي توضيح لبعض أهم هذه الاستخدامات :

1. **الانتقاء Selection :** "تستخدم في التنبؤ بالنجاح أو الفشل في الأداء الأكاديمي، أو المهني، ففي مواقف الانتقاء يتقدم عادةً عدداً من الأفراد يزيد عما هو مطلوب ، لذلك ينبغي اتخاذ قرار يتعلق بقبول أصلح الأفراد للالتحاق بدراسة أو مهنة معينة ويكون الغرض من الاختبار المساعدة في اتخاذ قرار انتقاء الأفراد الذين يكون احتمال نجاحهم كبيراً".

2. **التسكين Placement :** "بعد انتقاء الأفراد لدراسة أو مهنة معينة ، ينبغي تسكينهم في أحد البرامج أو الأقسام المناسبة، والهدف إحداث مزاوجة أمثل بين سمات الفرد التي تم قياسها ، ومتطلبات الدراسة أو العمل" (علاّم،2006: 41).

3. **التشخيص والعلاج Diagnosis and Remedial :**" في بعض الأحيان التدريس لفرد أو مجموعة من الأفراد غير ذي جدوى . إذ ربما يكون التعلم بحاجة إلى مادة تعليمية معينة ، أو أن يتم تدريسه بأسلوب ما ، وتحت ظروف ومعطيات معينة . "، وربما تعمل أدوات التشخيص إلى الإجابة عن سؤال مفاده : ما الأنشطة التعليمية التي تجعل أداء المتعلم أفضل ، أما قرار العلاج فيكون فعالاً إذا ما تمت عملية التشخيص بفاعلية كافية" (النبهان ، 2004 : 408) .

4. **الإرشاد والتوجيه Counseling and Guidance :**" إنَّ القرار لوضع الطالب في صف أو برنامج لهو قرار موجّه بشكل رئيس نحو اختيار أفضل فرص التعلّم لذلك الطالب وتوجيهه إلى خبرات يمكن أن تكون مجزية وفعّالة.. وتتعلق هذه الاختبارات من جهة بمدى تقدّم الطالب أكاديمياً بمستوى مقبول ، ومن جهة أخرى بخططه المستقبلية "(ثورندايك وهيجن ، 1989 : 171) .

5. **البحث التربوي والتقني :**" تُستخدم الاختبارات في التحقق من صحة الفروض، فعلماء النفس والتربية يستعملون الاختبارات في قياس متغيرات تتضمنها فروضهم البحثية من أجل التوصل إلى نظريات تفسر السلوك الإنساني ، وبذلك تُعدُّ أدوات علمية، ربما لا يكون لها تطبيقات عملية مباشرة في حياة الأفراد وبالطبع ليست هذه هي الاستخدامات الوحيدة للاختبارات والمقاييس التربوية والنفسية ، إذ أن هناك الكثير من الاستخدامات التي تناسب مختلف الأغراض ، استخدامه في المجالات التربوية يُعدُّ أكثر الاستخدامات انتشاراً في الدول المختلفة، فيساعد في تحسين وإثراء تعلم الطلاب، وتقويم تحصيلهم، ومتابعة تقدمهم الدراسي وتقديم تغذية راجعة للمعلم والطالب، وهي مصدر لزيادة دافعية الطلاب، وتطوير البرامج والمناهج التعليمية ، وتقويم مختلف مكونات المنظومة التربوية "(علاّم ، 2006 : 42-43).

طرق تصحيح الاختبار الصواب والخطأ المتعدد:

وقد أشار تساي وسوين (Tsain & Sun، 1993) إلى ست طرق لتقدير درجات فقرات الصواب والخطأ المتعدد، تعتمد على عدد البدائل الكلي في الفقرة وعدد البدائل التي استجاب لها المفحوص بطريقة صحيحة وعدد البدائل التي قد يتركها المستجيب بدون إجابة، ويمكن تطبيق هذه الطرق على الفقرات ذات البدائل الأربعة كما يلي:

- طريقة الاستجابات المتعددة Multiple Response Method أو ما يطلق عليها طريقة المعرفة التامة، وفيها تكون الفقرة ثنائية الاستجابة، حيث يحصل المستجيب على الدرجة (1) إذا كانت استجاباته لكل البدائل صحيحة ، ويحصل على الدرجة (صفر) إذا أخفق في استجابته لبديل أو اكثر من بديل.

- طريقة مراعاة وجود ثلاثة بدائل صحيحة Count for 3 Options Correct وفيها يحصل المستجيب على الدرجة (1) إذا أجاب عن كل البدائل بطريقة صحيحة، وعلى الدرجة (0,5) اذا كانت استجاب لثلاثة بدائل فقط بطريقة صحيحة، ويحصل على الدرجة (صفر) فيما عدا ذلك وبالتالي تكون الدرجات (1، 0.5 ، صفر).

- طريقة مراعاة وجود بديلين صحيحين Count for 2 Options Correct وفيها يحصل المستجيب على الدرجة (1) إذا أجاب عن كل البدائل بطريقة صحيحة، وعلى ثلثي درجة إذا استجاب لثلاثة بدائل فقط بطريقة صحيحة، ويحصل على ثلث درجة إذا استجاب البديلين فقط بطريقة صحيحة، ويحصل على الدرجة (صفر) فيما عدا ذلك، وبالتالي تكون الدرجات (1، ثلثين، ثلث، صفر).

- طريقة الصواب والخطأ المتعدد Multiple True - False Method : وفيها يحصل المستجيب على(0,25) درجة عن كل بديل يجيب عنه بشكل صحيح، وبالتالي تكون الدرجات (1، 0,75، 0,5، 0,25، صفر).

- طريقة تصحيح أثر التخمين Correction for Guessing Method وفيها يحصل المستجيب على الدرجة وفق المعادلة (1)

-

عدد البدائل الصحيحة - (عدد البدائل - عدد البدائل الصحيحة)
الدرجة= ———————————————
عدد البدائل

– طريقة السماح بترك البدائل غير المعروفة Let - Omet Method: وفيها يحصل المستجيب على الدرجة وفق المعادلة (2)

عدد البدائل الصحيحة - (عدد البدائل - عدد البدائل الصحيحة - عدد البدائل المتروكة)
الدرجة= ———————————————
عدد البدائل

ويلاحظ أن طريقة تقدير الدرجات الأولى تتطلب معرفة تامة بالمحتوى حيث يستطيع الطالب الحصول على الدرجة عند إجابته على جميع البدائل بشكل صحيح وإلا سوف يحصل على الدرجة (صفر)، وتتدرج الطرق الثانية والثالثة والرابعة في مراعاتها للمعرفة الجزئية لدى المستجيب من الأقل مراعاة لها إلى الأكثر مراعاة لها والطريقة الخامسة تستخدم لتصحيح أثر التخمين والطريقة السادسة تسمح للمستجيب بترك بديل أو أكثر بدون علامة، وكلتا الطريقتين الخامسة والسادسة تحتاجان إلى تعليمات خاصة عند تطبيق الاختبار على المستجيبين(عبد الوهاب ،2020: 3).

النظرية الكلاسيكية :

" شهدت حركة القياس النفسي تطورات مهمة في النصف الثاني من القرن العشرين، اسهمت في بلورة مفاهيمه الاساسية وفي ظهور بعض النظريات التي حاولت تفسيره وفي ايجاد اساليب ووسائل علمية جعلته يتبوأ مركز الصدارة بين فروع علم النفس ومجالاته العلمية المختلفة" (مخائيل، 1997: 7-13) "فلم يعد مقتصراً على إعداد الاختبارات أو المقاييس النفسية أو بنائها بل اخذ يهتم بتطوير المقاييس والاختبارات السابقة بإيجاد مؤشرات أو خصائص سيكومترية جديدة لها ودراسة تأثير بعض المتغيرات للوصول إلى خصائص سيكو مترية جيدة ومناسبة للاختبارات أو المقاييس النفسية التي تحققت دقتها في قياس ما أعدت لقياسه "(أبو علام، 1989: 7)."والنظرية الكلاسيكية تنطلق من فرض أساسي في بناء الاختبارات والمقاييس وتحليل فقراتها من كون توزيع درجات الأفراد في السمة أو الخصيصة التي يقيسها الاختبار تتوزع توزيعاً اعتدالياً (Normal Curve)، والذي يتأثر بطبيعة خصائص عينة الأفراد وخصائص عينة الفقرات "(Brown,1983:118)، "وتستند هذه النظرية على مبدأ الفروق الفردية ، لذا تقارن درجة الفرد فيها بدرجات المجموعات التي ينتمي إليها" (Crocker & Algian,1986:114) .

وتستند هذه النظرية إلى مجموعة من المسلمات أهمها:

1. أن أداء الفرد يمكن قياسه وتقديره ، أي يمكن تحويل أي نمط من الأنماط السلوكية للفرد من صيغة وصفية إلى صيغة كمية ، على الرغم من تداخله مع أنماطه السلوكية الأخرى من جهة ، ومع الظروف والعوامل المؤثرة في موقف القياس وعملية القياس من جهة أخرى.
2. أن أداء الفرد دالة خصائصه ، أي أن كل أداء أو سلوك يصدر عن خاصية أو أكثر ، كما أن الخاصية الواحدة يمكن أن تعطي أكثر من أداء ، وبهذا نجد أن هناك علاقة بين الخاصية والأداء ، وان تعقيد هذه العلاقة يؤثر في الأداة المستخدمة في القياس من حيث البناء والتكوين والدلالة والتفسير .
3. الخاصية والأداء والعلاقة بينهما تختلف من فرد إلى أخر(عبد الرحمن، 1983: 81 – 85).

يحصل عليها كل منهم على الاختبار ، ويتم تصنيف الأفراد بحسب مركزهم النسبي بين أقرانهم في القدرات أو السمات المختلفة (علام، 2001: 14 – 16) .

معالم الفقرات وفقاً لنظرية القياس الكلاسيكية **items features According of Traditional theory:**

اولاً : صعوبة المفردات Difficulty Items :

"ترى ضرورة ان تكون صعوبة فقرات الاختبار ممتدة على مدى واسع، اذ ان الفقرات السهلة تشجع الطلبة على الاجابة وتزيد الثقة بأنفسهم" (Anastasi , 1976: 152).

ثانياً : تمييز الفقرات Item Discrimination (D):

ترى" النظرية الكلاسيكية انا لفقرة الجيدة هي التي تميز او تحدد الفرق بين المستجيبين ذوي القدرة العالية وذوي القدرة الواطئة لذلك يتم اللجوء الى المجموعتين المتطرفتين في الدرجة الكلية –Extreme (groups)، اذ تمثل الدرجة الكلية مجموع الفقرات في الاختبار بافتراض ان الاختبار هو عينة للنطاق السلوك ومن ثم تتم المقارنة بين عدد المفحوصين الذين اجابوا عن الفقرة بصورة صحيحة في المجموعتين العليا والدنيا، وان الفقرة ذات التمييز المثالي الفعال هي تلك التي يجيب عنها بصورة صحيحة كل الافراد في المجموعة العليا، في حين يجيب عنها بصورة مخطئة كل الافراد في المجموعة الدنيا".

.(Brennan,1972:289-290)(Freeman,1962:23)

ثالثاً: ثبات المفردة Item Reliability (D) :

ان "دليل ثبات الفقرة هو قياس اسهام الفقرات في تباين الاختبار الكلي ومن ثم في ثباته ، أي يفحص ثبات الفقرة لتعرُف على هل كان ثبات الاختبار سيزيد عند اضافة تلك الفقرة او لا، وهذا يتم بفحص هل كانت الفقرة تقيس العامل نفسه الذي تقيسه الفقرات الاخرى او لا؟ ان الارتباط الموجب بين الفقره المطلوب فحص ثباتها والفقرات الاخرى في الاختبار ستشير الى هل كانت تلك الفقره ستزيد ثبات الاختبار، وللحصول على نتيجة صحيحة ينبغي استبعاد درجة تلك المفردة من الدرجة الكلية عندما يحسب معامل الارتباط " .(Magnusson,1967:214-215)

رابعا : صدق الفقرات Items Validity :

"الرغم من ان دليل التمييز يعد احد مؤشرات صدق المفردة، الا ان النظرية الكلاسيكية تنظر لارتباط المفردة بمحك خارجي او داخلي بوصفه الدليل الانسب لصدق المفردة، وتعد طريقة الاتساق الداخلي من أدق الوسائل لحساب صدق الفقرات في قياس المفهوم، فهي تقدم اختباراً متجانساً بحيث ان كل الفقرة تقيس البعد السلوكي الذي يقيسه الاختبار ككل، كما أنها تبرز الترابط الموجود بين فقرات الاختبار ككل" (عبد الرحمن، 1983 : 414).

الخصائص القياسية وفقاً للنظرية الكلاسيكية Psychometric Properties according to classical theory:

اولاً: صدق الاختبار :

"يكاد يتفق المتخصصون في القياس النفسي على إن الصدق يعدّ من أهم الخصائص السيكو مترية التي يجب توافرها في الاختبار التحصيلي لكونه يؤثر في قدرة الاختبار في قياس ما وضع من أجل قياسه، لذلك يمكن أن تندرج جميع الخصائص السيكو مترية الأخرى تحت خاصية الصدق" (Kline,1979:1). (داود & نور 2020، 224)

ثانياً: ثبات الاختبار :

"يشير ثبات الاختبار إلى دقة واتساق درجاته في قياس ما يجب قياسه ، ويقصد ايضا أن يكون متسقا في تقدير الدرجة الحقيقية للفرد في السمة التي يقيسها وذلك بالا يظهر نتائج متناقضة عند تكرار استخدامه على الفرد نفسه ولعدة مرات" (عودة وملكاوي، 1987: 194).

الدراسات السابقة

الصمادي(2012):(أثر طريقة اختبار الصواب – الخطأ المتعدد على الخصائص السيكومترية للاختبار وفقراته) هدفت هذه الدراسة للكشف عن أثر طريقة تصحيح اختبار الصواب والخطأ المتعدد، في معاملات الصدق والثبات للاختبار، ومعاملات الصعوبة والتمييز لفقرات الاختبار وكذلك أثرها في تحصيل الطلبة ولتحقيق هذه الأهداف تم بناء اختبار تحصيلي لمساق القياس والتقويم، يتكون من (٢٥) فقرة صواب وخطأ متعدد لكل منها أربعة بدائل. وطبق الاختبار على (١٢٢) طالب وطالبة في جامعة أربد الأهلية، تم تصحيح الاختبار باستخدام أربع طرق مختلفة للتصحيح متدرجة في مراعاتها للمعرفة الجزئية، وباستخدام برمجية (SPSS) تم حساب معاملي صدق للاختبار، ومعامل الثبات ألفا كرونباخ، وكذلك حسبت معاملات الصعوبة والتمييز لفقرات الاختبار ومتوسطات التحصيل، ولكل طريقة من طرق التصحيح المختلفة. وأظهرت نتائج هذه الدراسة عدم وجود فروق ذات دلالة إحصائية بين معاملات الصدق والثبات والتمييز، في حين أشارت النتائج لوجود فروق ذات دلالة إحصائية في معاملات صعوبة

الفقرات وفي تحصيل الطلبة، حيث كانت طريقة التصحيح الرابعة - الأكثر مراعاة للمعرفة الجزئية - هي الأكثر فاعلية بالنسبة لاختبارات التصحيح محكية المرجع ، وتلتها طريقة التصحيح الثالثة ثم الثانية فالأولى. وهذا الترتيب يظهر أن طريقة التصحيح الرابعة هي الأسهل حيث تعطي العلامة الأعلى(الصمادي ،2012 : 282-306).

دراسة :عبد الوهاب(2020) (أثر طريقة تقدير الدرجات ونموذج الاستجابة في دقة تقدير بارامترات الافراد والمفردات لاختبار صواب والخطأ متعدد) هدف البحث إلى معرفة أثر طرق تقدير الدرجات على مفردات الصواب والخطأ المتعدد ونوع نموذج الاستجابة للمفردة المستخدم التقدير الجزئي والاستجابات المتدرجة) والتفاعل بينهما على دقة تقدير بارامترات الأفراد والمفردات وفي سبيل ذلك تم بناء اختبار تحصيلي في مقرر علم النفس التربوي لطلاب كلية التربية بجامعة أم القرى (411) طالبا بحيث يتكون من (35) مفردة من نوع الصواب والخطأ المتعدد، وتم تقدير الأخطاء المعيارية في تقدير قدرات الطلاب على هذا الاختبار وبارامترات الصعوبة وفق أربع طرق مختلفة لتقدير الدرجات وباستخدام نموذجي التقدير الجزئي والاستجابات المتدرجة، كما تم تقدير الأخطاء المعيارية في تقدير بارامترات تمييز المفردات باستخدام نموذج الاستجابات المتدرجة، وقد أسفرت النتائج عن وجود أثر لكل من طريقة تقدير الدرجات ونوع النموذج المستخدم والتفاعل بينهما في دقة تقدير بارامترات صعوبة المفردات وبارامترات القدرة للطلاب، كما وجد أثر لطريقة تقدير الدرجات في دقة تقدير بارامترات تمييز المفردات وفق نموذج الاستجابات المتدرجة، وقد أوصى البحث بأفضلية الاعتماد على الطريقة التي تراعي المعرفة الجزئية بأكبر قدر ممكن في تقدير الدرجات على المفردات، وكذلك أفضلية استخدام نموذج الاستجابات المندرجة في تدريج هذه المفردات **(عبد الوهاب ،2020 :1-22).**

الفصل الثالث : منهجية البحث وإجراءاته:

منهجية البحث:

هناك عدد من المناهج في اجراء البحوث في العلوم الانسانية فطبيعة البحث وموضوع الدراسة والهدف منها هو الذي يحدد نوع البيانات المراد جمعها وطبيعة المنهج المستعمل في اجراء البحث، وبما ان البحث الحالي يهف الى بناء اختبار تحصيلي الصواب والخطأ المتعدد في مادة منهج البحث العلمي لطلبة الصف الثالث لكليات التربية في جامعات بغداد، فان المنهج الملائم لموضوع البحث هو المنهج الوصفي المقارن لا اسباب اي (دراسات السببية المقارنة).

مجتمع البحث :

يتكون مجتمع البحث الحالي من طلب الجامعات بغداد كليات التربية الصف الثالث الذين درسوا مادة منهج البحث العلمي (وفق مفردات الهيئة القطاعية لكليات التربية الجامعات العراقية) حيث بلغ المجموع الكلي للطلبة (4438) طالب وطالبة وكان عدد الذكور (1420)بنسبة (32 %)والاناث (3018) بنسبة(68 %).

عينة البحث :

العينة هي جزء من المجتمع المستهدف لدراسة ولا جل الوصول الى نتائج دقيقة يمكن الوثوق بها ،يتطلب اخذ عدد من العينات بحسب الاجراء المتبع لتحقيق اهداف البحث .

اداة البحث :

ولتحقيق اهداف البحث يتطلب بناء اختبار تحصيلي الصواب والخطأ المتعدد لمادة منهج البحث العلمي وتبعت الباحثة الخطوات بناء الاختبار التحصيلي التالية :

- **تحديد المحتوى المادة الدراسية :**

من خلال اطلاع الباحثة على مفردات مادة منهج البحث العلمي المثبتة من قبل الهيئة القطاعية لكليات التربية الجامعات العراقية في تدريس المقرر تم تحديد مفرادات المادة .

- **صياغة الاهداف السلوكية**

يؤكد التربويون اهمية تحديد الاهداف السلوكية في بناء الاختبارات التحصيلية والتي يتم قياسها بالاختبار من الخطوات المهمة في البناء وكذلك التأكيد على صياغة الاهداف ان تكون واضحة ومحددة وقابلة للقياس والملاحظة (شحاتة ،1998: 61)

حيث صاغت الباحثة (25)هدف سلوكي موزع على الاهداف السلوكية لتصنيف بلوم (تذكر ،فهم ،تطبيق)

- اعداد جدول مواصفات الاختبار التحصيلي :

تم تحديد (25) فقرة ،وكل فقرة تقيس هدف سلوكي واحد لاختبار تحصيلي الصواب والخطأ المتعدد وفي ضوء الاهمية النسبة .

- **صياغة فقرات اختبار الصواب والخطأ المتعدد :**

قامت الباحثة بصياغة (25) فقرة موضوعية من نوع الصواب والخطأ المتعدد ،حيث هذا النوع من الاختبار يشبه اختبار الاختيار من متعدد حيث يحتوي على اصل الفقرة ولكل فقرة اربع بدائل ، وتم صياغة البدائل الاجابة حيث احتوى الاختبار على فقرات بعضها تحتوي على بديل واحد فقط صحيح والبدائل الاخرى خطأ وبعظها تحتوي على بديلين صح والبديلين الاخريين خطأ وبعض الفقرات تحوي على ثلاث بدائل صحيحة وواحدة خطأ وبعضها كلها صحيحة وبعضها كلها خطأ.

- **تعليمات الاختبار :**

تم وضع تعليمات في مقدمة الاختبار وهذه التعليمات الخاصة بكيفية الاجابة على بدائل فقرات الاختبار جميها وعدم ترك فقرة او بديل بدون اجابة ،حيث وضحت الباحثة تعليمات لكيفية الاجابة على كل البدائل من خلال وضع مثال توضيحي لكيفية الاجابة على فقرات الاختبار وتم احتساب وقت الاجابة على فقرات الاختبار بلغ متوسط وقت الاجابة (28) دقيقة .

- **ورقة الاجابة :**

تم تصميم ورقه خاصة منفصلة عن كراس الاختبار وضعت تسلسل الفقرات وبدائل الاجابة الأربعة لكل فقرة من فقرات الاختبار حيث يطلب من المفحوصين وضع لكل بديل علامة (√) أو علامة (×) لجميع البدائل لكل فقرة من فقرات الاختبار .

- تصحيح اجابات الاختبار وفق الطرق التصحيح الاربعة التالية :

أ- اعطاء درجة (1) لكل فقرة اجاب عنها الطالب عن كل البدائل (أ،ب،ج، د) بشكل صحيح، وصفر اذا اجاب بشكل غير صحيح على اي بديل من البدائل الاربعة (1، 0) .

ب- درجة (1) اذا اجاب على كل البدائل بشكل صحيح اعطاء نصف درجة اذا اجاب على بديلين بشكل صحيح وصفر اذا اجاب عن بديل واحد صح او جميع البدائل بشكل غير صحيح (1، 0,5، 0).

ت- اعطاء ثلث الدرجة اذا اجاب الطلب على بديلين بشكل صحيح وثلثين اذا اجاب على ثلاث بدائل بشكل صحيح ودرجة (1) اذا اجاب على اربع بدائل بشكل صحيح وصفر اذا اجاب على بديل واحد صحيح واذا اجاب على البدائل كلها بشكل خطأ.(0,333، 0,666، 1، 0)

ث- اعطاء ربع درجة اذا اجاب الطلب على البديل واحد ونصف اذا بديلين وثلاث ارباع اذا اجاب على البدائل ثلاثة ودرجة (1) اذا اجاب على اربع البدائل بشكل صحيح وصفر اذا لم يجيب على اي بديل بشكل صحيح .(0,25 ، 0,5، 0,75، 1)

التحليل الاحصائي لفقرات اختبار التحصيل في مادة منهج البحث العلمي بالطريقة الاولى (الاعتيادية المتمثلة بأ عطاء (واحد) اذا اجاب عن البدائل جميعها بشكل صحيح وصفر اذ لم يجب بشكل صحيح عن جميع البدائل):

- **عينة التحليل الاحصائي**

للتحقق من الخصائص السيكو مترية لفقرات الاختبار قامت الباحثة بتطبيق الاختبار على عينة مكونة من (400) طالب وطالبة ، قامت الباحثة بحساب الخصائص القياسية وفقاً للخطوات الاتية:-

- **معامل صعوبة الفقرة :**

تم حساب معامل الصعوبة لكل فقرة من فقرات الاختبار باستخدام معادلة (معامل الصعوبة) الخاصة باختبار الصواب والخطأ المتعدد وكالاتي:

1- ترتيب الدرجات التي حصل عليها الطلاب من اعلى درجة الى اوطأ درجة .

2- تم اختيار نسبة 27% العليا والدنيا من الدرجات لتمثل المجموعتين المتطرفتين ، فقد شملت المجموعتان (216) طالب وطالبة ، بحيث ان كل مجموعة ضمت (108) طالب وطالبة.

3- استخرج عدد الطلاب الذين اجابوا اجابة خاطئة في كل من المجموعتين العليا والدنيا عن كل فقرة من فقرات الاختبار .

4- *جمع الاجابات الخاطئة للمجموعتين وتقسيمها على العدد الكلي للمجموعتين* بمعادلة صعوبة الفقرات للأسئلة الموضوعية اذ وجدتها تراوحت بين(0.333) و(0.611) ، وتعد الفقرات جيدة إذا تراوح معامل صعوبتها بين (0.20-0.80) (الظاهر وآخرون، 2002 :129). وبهذا تعد فقرات الاختبار مقبولة ومعامل صعوبتها مناسباً من ناحية هذا المؤشر الإحصائي كما موضحة بالجدول (1).

معامل التمييز للفقرات :

بعد حساب معامل التمييز لكل فقرة من فقرات الاختبار الموضوعية وجدت انها انحصرت بين (0,333-0,593) والجدول (1) يوضح ذلك، اذ تعد الفقرة مقبولة اذا كان معامل تمييزها يزيد على (0,30) (الظاهر وآخرون ،129: 1999)، وبذلك تعد فقرات الاختبار جميعها مقبولة .

جدول (1)

الاجابات الصحيحة والاجابات الخاطئة للمجموعتين العليا والدنيا ومعامل الصعوبة والسهولة والقوة التمييزية لفقرات اختبار الصواب والخطأ المتعدد

الفقرة	عدد الاجابات في المجموعة العليا		عدد الاجابات في المجموعة الدنيا		عدد الاجابات الصحيحة	عدد الاجابات الخاطئة	معامل السهولة	معامل الصعوبة	معامل التمييز
	الصحيحة	الخاطئة	الصحيحة	الخاطئة					
1	96	12	48	60	144	72	0.667	0.333	0.444
2	64	44	28	80	92	124	0.426	0.574	0.333
3	64	44	24	84	88	128	0.407	0.593	0.370
4	60	48	24	84	84	132	0.389	0.611	0.333
5	68	40	24	84	92	124	0.426	0.574	0.407
6	76	32	36	72	112	104	0.519	0.481	0.370
7	80	28	44	64	124	92	0.574	0.426	0.333
8	96	12	36	72	132	84	0.611	0.389	0.556
9	92	16	32	76	124	92	0.574	0.426	0.556
10	76	32	40	68	116	100	0.537	0.463	0.333
11	80	28	40	68	120	96	0.556	0.444	0.370
12	80	28	32	76	112	104	0.519	0.481	0.444
13	80	28	40	68	120	96	0.556	0.444	0.370
14	84	24	40	68	124	92	0.574	0.426	0.407
15	64	44	24	84	88	128	0.407	0.593	0.370
16	68	40	32	76	100	116	0.463	0.537	0.333
17	76	32	40	68	116	100	0.537	0.463	0.333

الفقرة	عدد الاجابات في المجموعة العليا		عدد الاجابات في المجموعة الدنيا		عدد الاجابات الصحيحة	عدد الاجابات الخاطئة	معامل السهولة	معامل الصعوبة	معامل التمييز
	الصحيحة	الخاطئة	الصحيحة	الخاطئة					
18	96	12	48	60	144	72	0.667	0.333	0.444
19	88	20	28	80	116	100	0.537	0.463	0.556
20	80	28	36	72	116	100	0.537	0.463	0.407
21	68	40	28	80	96	120	0.444	0.556	0.370
22	80	28	24	84	104	112	0.481	0.519	0.519
23	72	36	36	72	108	108	0.500	0.500	0.333
24	84	24	32	76	116	100	0.537	0.463	0.481
25	88	20	24	84	112	104	0.519	0.481	0.593

- **علاقة الفقرة بالدرجة الكلية لاختبار التحصيلي الصواب والخطأ المتعدد**

علاقة درجة كل فقرة من الفقرات بالدرجة الكلية للاختبار. تم بأستعمال معامل ارتباط بوينت بايسيريال ،تبين أن جميعها دالة احصائيا عند مقارنتها بالقيمة الحرجة لمعامل الارتباط البالغة (0.098) عند مستوى دلالة (0.05) ودرجة حرية (398) ، والجدول (2) يوضح ذلك.

جدول (2)

معاملات ارتباط بوينت بايسيريال بين درجة كل فقرة من الفقرات والدرجة الكلية للاختبار

ت	معامل الارتباط	ت	معامل الارتباط	ت	معامل الارتباط
1	0.389	11	0.345	21	0.447
2	0.521	12	0.456	22	0.476
3	0.369	13	0.439	23	0.409
4	0.447	14	0.541	24	0.511
5	0.410	15	0.447	25	0.456
6	0.483	16	0.408		
7	0.509	17	0.396		
8	0.467	18	0.431		
9	0.422	19	0.433		
10	0.401	20	0.489		

- **الخصائص السيكو مترية لاختبار التحصيلي منهج البحث العلمي :**

اولاً: صدق الاختبار Test Validity

تم التأكد من صدق المقياس كالاتي:

أ. الصدق الظاهري Face Validity

من خلال عرض الاختبار على مجموعة من المحكمين لتبيان مدى صلاحية الفقرات الاختبارية تم التحقق من الصدق الظاهري ، وقد كانت جميع الفقرات صالحة ظاهرياً في قياس ما اعدت لقياسه في ضوء تقديرات المحكمين.

ب.صدق المحتوى Content Validity

من خلال أعداد الخارطة الاختبارية ، يتم تمثيل الفقرات الاختبارية للمحتوى العلمي وتحقق الباحثة من صدق المحتوى.

ج. صدق البناء:

لقد تحقق هذا النوع من الصدق كالآتي:
أ- إيجاد معامل السهولة والصعوبة والقوة التمييزية لفقرات اختبار التحصيل بواسطة أسلوب المجموعتين المتطرفتين حيث تبين ان جميع فقرات الاختبار متوسطة السهولة والصعوبة وذات تمييز مقبول والجدول (1) يوضح ذلك.
ب- الارتباط بين الفقرة والدرجة الكلية للاختبار .

ثانياً:- ثبات الاختبار :
تم التحقق من ثبات الاختبار :
▪**معادلة الفا كرونباخ:**
"لاستخراج الثبات بهذه الطريقة، طبقت معادلة (الفا كرونباخ) على درجات افراد العينة البالغ عددهم (400) طالب وطالبة ، فكانت قيمة معامل ثبات الاختبار (0,763) ، وبذلك تُـعد قيمة جيدة ومناسبة لذا يعُد الاختبار ثابتـاً، "حيث إن الاختبارات غير مقننة تـُعد جيدة اذ ما بلغ معامل ثباتها (0,67) فما فوق" (عودة والخليلي، 1988: 146).
وبذلك اصبح الاختبار بصورته النهائية مؤلفا من (25) فقرة اختبار الصواب والخطأ المتعدد جاهزاً للتطبيق .

- الوسائل الاحصائية:
1. استعانت الباحثة بالحقيبة الاحصائية للعلوم الاجتماعية (SPSS) لاستخراج :
- معادلة ألفا كرونباخ.
- تحليل التباين الاحادي.
- اختبار شيفيه للمقارنات البعدية.
- الاختبار الزائي للمقارنة بين معاملي ارتباط.
ثانياً: برنامج Microsoft Office Excel 2010 وذلك لاستخراج :
▪**معامل صعوبة وسهولة الفقرة الموضوعية**: استخدم في ايجاد سهولة وصعوبة الفقرات لاختبار التحصيل

$$D = \frac{N_1 + N_2}{N}$$

D: معامل الصعوبة
N: عدد افراد المجموعتين العليا والدنيا
N_1: عدد الافراد الذين اجابوا اجابة خاطئة من المجموعة العليا
N_2: عدد الافراد الذين اجابوا اجابة خاطئة من المجموعة الدنيا
▪**الصعوبة = 1- معامل السهولة.**
▪**معادلة تمييز الفقرات الموضوعية :** استخدم في ايجاد تمييز الفقرات لاختبار التحصيل .

$$D_E = \frac{N_1 - N_2}{\frac{N}{2}}$$

حيث أن D_E : معامل التمييز
N: عدد افراد المجموعتين العليا والدنيا
N_1: عدد الافراد الذين اجابوا اجابة صحيحة من المجموعة العليا.
N_2: عدد الافراد الذين اجابوا أجابة صحيحة من المجموعة الدنيا.
3 معامل الارتباط بوينت بايسيريال :
من اجل معرفة علاقة الفقرة بالدرجة الكلية لاختبار التحصيل:

$$rpb = \frac{\bar{X}.1 - \bar{X}.0}{Sx} \sqrt{\frac{n_0 * n_1}{n * (n-1)}}$$

(Rob) معامل الارتباط .
($\bar{X}.0$) يمثل الوسط الحسابي للدرجات لاختبار الطلاب الذين أجابوا إجابة خاطئة على تلك الفقرة
($\bar{X}.1$) يمثل الوسط الحسابي للدرجات لاختبار الطلاب الذين أجابوا إجابة صحيحة على تلك الفقرة
(n0) يمثل أعداد العينة الذين أجابوا على الفقرة تلك إجابة خاطئة .
(n1) يمثل أعداد العينة الذين أجابوا على الفقرة تلك إجابة صحيحة .
(SX) يمثل الانحراف المعياري للكل إفراد العينة .

الفصل الرابع :

عرض النتائج وتفسيرها

عرض النتائج الفرضيات:

1- الهدف الاول : بناء اختبار تحصيلي الصواب والخطأ المتعدد لمادة منهج البحث العلمي للصف الثالث في جامعات بغداد كليات التربية .

ولتحقيق هذا الهدف قامت الباحثة بجميع الاجراءات والخطوات العلمية اللازمة لبناء الاختبار وكما موضحة في الفصل الثالث ،وتبن ان جميع الفقرات والاختبار الكلي يتمتع بخصائص قياسية جيدة ،وتم الابقاء على جميع الفقرات ،وتم تحقيق الهدف .

2- الهدف الثاني : التعرف على دلالة الفروق في الخصائص القياسية للاختبار التحصيلي في مادة منهج البحث العلمي باختلاف طرق تصحيح اختبار الصواب والخطأ المتعدد (ط1، ط2، ط3، ط4) ويشتق من هذا الهدف الفرضيات التالية :

- الفرضية الاولى: لا توجد فروق ذات دلالة احصائية بين معاملات صعوبة الفقرات المحسوبة للطرائق الاربعة في تصحيح اختبار الصواب والخطأ (ط1 – ط2 – ط3 – ط4)، ولمعرفة دلالة الفرق بين الطرق الاربعة معاملات صعوبة الفقرات ، تم اعتماد تحليل التباين الأحادي (One Way ANOVA)، وباستعمال الحقيبة الإحصائية (SPSS)، فكانت النتائج كما موضحة في الجدول (3).

الجدول (3) الأوساط الحسابية والانحرافات المعيارية لقيم الصعوبة تبعا لطرق تصحيح اختبار الصواب والخطأ المتعدد (ط1 – ط2 – ط3 – ط4)

الطريقة	المتوسط الحسابي لقيم الصعوبة	الانحراف المعياري لقيم الصعوبة
الاولى	0.481	0.075
الثانية	0.469	0.068
الثالثة	0.496	0.057
الرابعة	0.484	0.058

يلاحظ من البيانات الواردة في الجدول رقم (3)أن هناك فروقاً في المتوسطات الحسابية لقيم الصعوبة ولتحديد فيما إذا كانت هذه الاختلافات ذات دلالة احصائية عند مستوى دلالة (0.05) تم استخدام تحليل التباين الاحادي والجدول رقم (4) يبين نتائج تحليل التباين الاحادي

جدول (4)

نتائج تحليل التباين الاحادي بين معاملات صعوبة الفقرات المحسوبة للطرائق الاربعة

مصدر التباين	مجموع المربعات S.S	درجة الحرية d.F	متوسط المربعات S.M	النسبة الفائية F		الحكم
				المحسوبة	الجدولية	

بين المجموعات B	0,009	3	0,003	0,706	2,68	غير دالة
داخل المجموعات W	0,408	96	0,004			
الكلي T	0,417	99				

وهذه النتيجة تشير الى عدم وجود فروق ذات دلالة احصائية في معاملات صعوبة الفقرات بين الطرق الاربعة اذ كانت القيمة الفائية المحسوبة والبالغة (0,706) هي اصغر من القيمة الفائية الجدولية والبالغة (2,68) بمستوى دلالة (0,05) وبدرجة حرية (3، 96).

- **الفرضية الثانية:** لا توجد فروق ذات دلالة احصائية بين معاملات تمييز الفقرات المحسوبة للطرائق الاربعة في تصحيح اختبار الصواب والخطأ المتعدد(ط1 – ط2 – ط3 – ط4)، ولمعرفة دلالة الفرق بين الطرق الاربعة في معاملات تمييز الفقرات ، تم اعتماد تحليل التباين الأحادي (One Way ANOVA)، وباستعمال الحقيبة الإحصائية (SPSS)، فكانت النتائج كما موضحة في الجدول (5).

الجدول (5) الأوساط الحسابية والانحرافات المعيارية لقيم التمييز تبعا لطريقة تصحيح اختبار الصواب والخطأ المتعدد (ط1 – ط2 – ط3 – ط4)

الطريقة	المتوسط الحسابي لقيم التمييز	الانحراف المعياري لقيم التمييز
الاولى	0.414	0.083
الثانية	0.426	0.065
الثالثة	0.476	0.066
الرابعة	0.532	0.051

يلاحظ من البيانات الواردة في الجدول رقم (5) أن هناك فروقاً في المتوسطات الحسابية لقيم التمييز ولتحديد فيما إذا كانت هذه الاختلافات ذات دلالة احصائية عند مستوى دلالة (0.05) تم استخدام تحليل التباين الاحادي والجدول رقم (6) يبين نتائج تحليل التباين الاحادي

جدول (6)

نتائج تحليل التباين الاحادي بين معاملات تمييز الفقرات المحسوبة للطرائق الاربعة

مصدر التباين	مجموع المربعات S.S	درجة الحرية d.F	متوسط المربعات S.M	النسبة الفائية F		الحكم
				المحسوبة	الجدولية	
بين المجموعات B	0,217	3	0,072	15,691	2,68	دالة
داخل المجموعات W	0,442	96	0,005			
الكلي T	0,658	99				

وهذه النتيجة تشير الى وجود فروق ذات دلالة احصائية في معاملات تمييز الفقرات بين الطرائق الاربعة، اذ كانت القيمة الفائية المحسوبة والبالغة (15,691) هي اكبر من القيمة الفائية الجدولية والبالغة

(2,68) بمستوى دلالة (0,05) وبدرجة حرية (3،96)، وللتعرف على مصدر الفرق تم استخدام اختبار شيفيه للمقارنات البعدية فكانت النتائج كما موضحة في الجدول (7).

جدول (7)

نتائج اختبار شيفيه للمقارنات البعدية لتعرف مصدر الفرق بين معاملات تمييز الفقرات للطرائق الاربعة

الاختبار	المقارنة الثنائية	الوسط الحسابي	الفرق الملاحظ	قيمة شيفيه الحرجة	اتجاه الدلالة
اختبار الصواب والخطأ المتعدد	ط1	0,414	0,012	0,055	غير دالة
	ط2	0,426			
	ط1	0,414	0,062		دالة لصالح ط3
	ط3	0,476			
	ط1	0,414	0,118		دالة لصالح ط4
	ط4	0,532			
	ط2	0,426	0,050		غير دالة
	ط3	0,476			
	ط2	0,426	0,106		دالة لصالح ط4
	ط4	0,532			
	ط3	0,476	0,056		دالة لصالح ط4
	ط4	0,532			

وفي ضوء النتائج الموضحة في الجدول اعلاه يتبين ان افضل طريقة هي الطريقة الرابعة اذ اظهرت النتائج وجود فروق في القوة التمييزية بين الطرق لصالح الطريقة الرابعة .

- **الفرضية الثالثة :** لا توجد فروق ذات دلالة احصائية بين معاملات ارتباط الفقرات بالدرجة الكلية للاختبار المحسوبة للطرائق الاربعة في تصحيح اختبار الصواب والخطأ المتعدد(ط1 – ط2 – ط3 – ط4)، ولمعرفة دلالة الفرق بين الطرق الاربعة في معاملات ارتباط الفقرات بالدرجة الكلية للاختبار ، تم اعتماد تحليل التباين الأحادي (One Way ANOVA)، وباستعمال الحقيبة الإحصائية (SPSS)، فكانت النتائج كما موضحة في الجدول (8).

الجدول (8) الأوساط الحسابية والانحرافات المعيارية لقيم ارتباط الفقرات بالدرجة الكلية للاختبار **تبعا لطريقة تصحيح اختبار الصواب والخطأ المتعدد**(ط1 – ط2 – ط3 – ط4)

الطريقة	المتوسط الحسابي لقيم ارتباط المفردات بالدرجة الكلية للاختبار	الانحراف المعياري لقيم ارتباط المفردات بالدرجة الكلية للاختبار
الاولى	0.437	0.077
الثانية	0.451	0.065
الثالثة	0.501	0.068
الرابعة	0.561	0.055

يلاحظ من البيانات الواردة في الجدول رقم (8)أن هناك فروقاً في المتوسطات الحسابية لقيم ارتباط الفقرات بالدرجة الكلية للاختبار ولتحديد فيما إذا كانت هذه الاختلافات ذات دلالة احصائية عند مستوى دلالة (0.05) تم استخدام تحليل التباين الاحادي والجدول رقم (9) يبين نتائج تحليل التباين الاحادي

جدول (9)

نتائج تحليل التباين الاحادي بين معاملات ارتباط الفقرات بالدرجة الكلية للاختبار الفقرات المحسوبة للطرائق الاربعة

مصدر التباين	مجموع المربعات S.S	درجة الحرية d.F	متوسط المربعات S.M	النسبة الفائية F		الحكم
				المحسوبة	الجدولية	
بين المجموعات B	0,245	3	0,082	18,241	2,68	دالة
داخل المجموعات W	0,429	96	0,004			
الكلي T	0,674	99				

وهذه النتيجة تشير الى وجود فروق ذات دلالة احصائية في معاملات **ارتباط الفقرات بالدرجة الكلية للاختبار** الفقرات بين الطرائق الاربعة، اذ كانت القيمة الفائية المحسوبة والبالغة (18,241) هي اكبر من القيمة الفائية الجدولية والبالغة (2,68) بمستوى دلالة (0,05) وبدرجة حرية (3،96)، وللتعرف على مصدر الفرق تم استخدام اختبار شيفيه للمقارنات البعدية فكانت النتائج كما موضحة في الجدول (10).

جدول (10)

نتائج اختبار شيفيه للمقارنات البعدية لتعرف مصدر الفرق بين معاملات ارتباط الفقرات بالدرجة الكلية للاختبار الفقرات للطرائق الاربعة

الاختبار	المقارنة الثنائية	الوسط الحسابي	الفرق الملاحظ	قيمة شيفيه الحرجة	اتجاه الدلالة
اختبار الصواب والخطأ المتعدد	ط1	0,437	0,014	0,051	غير دالة
	ط2	0,451			
	ط1	0,437	0,064		دالة لصالح ط3
	ط3	0,501			
	ط1	0,437	0,124		دالة لصالح ط4
	ط4	0,561			
	ط2	0,451	0,050		غير دالة
	ط3	0,501			
	ط2	0,451	0,110		دالة لصالح ط4
	ط4	0,561			
	ط3	0,501	0,060		دالة لصالح ط4
	ط4	0,561			

وفي ضوء النتائج الموضحة في الجدول اعلاه يتبين ان افضل طريقة هي الطريقة الرابعة اذ اظهرت النتائج وجود فروق في معاملات ارتباط الفقرات بالدرجة الكلية للاختبار بين الطرق لصالح الطريقة الرابعة .

- **الفرضية الرابعة :** لا توجد فروق ذات دلالة احصائية بين معاملات الثبات المحسوبة للطرائق الاربعة في تصحيح اختبار الصواب والخطأ المتعدد (ط1 – ط2 – ط3 – ط4)، للتحقق من هذه الفرضية تم استخدام الاختبار الزائي للمقارنة بين معاملي ارتباط (يعد معامل الثبات هو بمثابة معامل ارتباط)، فكانت النتائج كما موضحة في الجدول (11).

جدول (11)
نتائج الاختبار الزائي للمقارنة بين قيم معاملات الثبات المحسوبة للطرائق الاربعة

طريقة التصحيح	العدد	معامل الثبات	الدرجة المعيارية	القيمة الزائية		مستوى الدلالة	الحكم
				المحسوبة	الجدولية		
ط1	400	0,763	1,008	1,084	1,960	0,05	غير دالة
ط2	400	0,797	1,085				
ط1	400	0,763	1,008	2,760			دالة لصالح ط3
ط3	400	0,835	1,204				
ط1	400	0,763	1,008	5,183			دالة لصالح ط4
ط4	400	0,887	1,376				
ط2	400	0,797	1,085	1,676			غير دالة
ط3	400	0,835	1,204				
ط2	400	0,797	1,085	4,098			دالة لصالح ط4
ط4	400	0,887	1,376				
ط3	400	0,835	1,204	2,422			دالة لصالح ط4
ط4	400	0,887	1,376				

وفي ضوء النتائج الموضحة في الجدول اعلاه يتبين ان افضل طريقة هي الطريقة الرابعة اذ اظهرت النتائج وجود فروق في الثبات بين الطرق لصالح الطريقة الرابعة.

الاستنتاجات :

في ضوء النتائج التي توصلت الية الباحثة :

1- دقة الاختبار التحصيلي الذي بنى في البحث الحالي وذلك لانخفاض قيمة الانحراف المعياري وارتفاع معامل الثبات .

2- ومن خلال طرق المفاضلة في التصحيح الاختبار الصواب والخطأ المتعدد ومن خلال معرفة دلالة الفروق بين الاوساط الحسابية والانحراف المعياري لقيم الصعوبة والتمييز والدرجة الكلية وقيمة الثبات لطرق تصحيح اختبار الصواب والخطأ المتعدد، اتضحت ان الطريقة الرابعة كانت أفضل الطرق في تصحيح اختبار الصواب والخطأ المتعدد وتليها الطريقة الثالثة ثم الطريقة الثانية، ثم الطريقة الاولى. وهذه النتيجة تتطابق مع نتائج الدراسات السابقة دراسة (الصمادي، 2012).حيث تتطلب الطرق الاخرى معرفة اكثر من الطريقة الرابعة والتي يحصل الطالب على درجة اعلى وفق هذه الطريقة .

التوصيات :

1- نوصي الاخذ بالطريقة الرابعة افضل طرق تصحيح اختبار الصواب والخطأ المتعدد التي يتوقع ان يحصل الطالب على اعلا درجة من الطرق الاخرى وكذلك تهتم بالمعرفة الجزئية للطلبة عن الطرق الاخرى التي تتطلب معرفه اكثر .

2- نظرا لحداثة الاختبارات الصواب والخطأ المتعدد وتعدد طرق تصحيحها توصي الباحثة اجراء المزيد من الدراسات والبحوث للتحقيق من المزايا في بناء اختبارات التربوية

3- استعمال هذه الاختبارات لغرض معرفة قدرات الطلبة بشكل عام والطلبة المتأخرين والمتفوقين لغرض تقديم لهم المساعدة .

4- استعمال الاختبار في مجالات الارشاد والتوجيه النفسي والتربوي وجعلها ادوات تشخيصية لمستويات الطلبة ورفع مستواهم ضمن الوسط الطلابي .

المقترحات :

1- دراسة اثر حجم العينة وطول الاختبار في الخصائص السيكو مترية الاختبار الصواب والخطأ المتعدد .

2- اجراء دراسة حول المفاضلة بطرق تصحيح الاختبار الحالي وفق نظري القياس الحديثة (نظرية السمات الكامنة) ووفق نماذجها .

المصادر العربية

- أ. د. داود عبدالسلام صبري & ,نور علي عبدالله. (2020). اثر برنامج تعليمي على وفق نظرية الذكاء الثلاثي في تحصيل مادة علم النفس الفسيولوجي لدى طلبة كلية التربية .مجلة الفنون والأدب وعلوم الإنسانيات والاجتماع.223-211 ,(62) ,
- عودة، احمد سليمان (2000): القياس والتقويم في العملية التدريسية، ط4، الأردن، دار الأمل للنشر.
- ابو فوده، باسل خميس وبني يونس، نجاتي احمد (2012): الاختبارات التحصيلية، ط1، عمان، الاردن، دار المسيرة للنشر والتوزيع والطباعة.
- البغدادي، محمد رضا (1984) : الاهداف والاختبارات بين النظرية والتطبيق في المناهج وطرق التدريس، الكويت، مكتبة الفلاح.
- علام، صلاح الدين محمود (2006): الاختبارات والمقاييس التربوية والنفسية ، ط1 ، الأردن، دار الفكر للنشر والتوزيع .
- علام ،صلاح الدين محمود (2011):القياس والتقويم التربوي في العمليةالتدريسية،ط4 ،دار المسيرة للنشر والتوزع ،عمان .
- علام ،صلاح الدين محمود(2014):الاختبارات والمقاييس التربوية والنفسية ،ط4،عمان ،دار الفكر لطباعة والتوزيع .
- النبهان، موسى (2004): أساسيات القياس في العلوم السلوكية، عمّان، الأردن، دار الشروق للنشر والتوزيع.
- ثورندايك، روبرت وهيجن، اليزابيث (1989): القياس والتقويم في علم النفس والتربية، ترجمة عبد الله زيد الكيلاني، وعبد الرحمن عدس، عمّان، مركز الكتب الأردني.
- عبد الوهاب(2022): أثر طريقة تقدير الدرجات ونموذج الاستجابة للمفردة في دقة تقدير بارامترات الافراد والمفردات لاختبار صواب وخطا متعدد ، جامعه ام القرى ، كلية التربية.
- ابو جراد ، (2015):اثر اختلاف عدد البدائل الصحيحة في فقرات الصواب والخطأ المتعدد في دقة تقدير صعوبة الفقرات وقدرات الافراد وثبات الاختبار ،جامعه عين الشمس ، مجلة البحث العلمي في التربية المجلد(16)الجزء الثالث العدد(3) مصر ،ص 1-20 .
- عبد الوهاب ، محمد محمود(2020):اثر فعالية بدائل مفردات الصواب والخطأ المتعدد على معالم المفردات والافراد في ضوء نظرية القياس التقليدية ونموذج الاستجابات المتدرجة ،جامعه الملك سعود ، كلية التربية.
- الصمادي، مروان صالح (2012) : اثر طريقة تصحيح اختبار الصواب والخطأ المتعدد على الخصائص السيكو مترية للاختبار وفقراته ،جامعه البحرين ،مجلة العلوم التربوية والنفسية ،مجلد (13) العدد(3) ص282- 306.
- الظاهر ،زكريا محمد واخرون (1999): مبادىء القياس والتقويم في التربية ،عمان ،الاردن ،دار الثقافة للنشر والتوزيع .
- عودة ،احمد سليمان ،والخليلي ،خليل يوسف (1998): الاحصاء للباحث في التربية والعلوم الانسانية ،دار الفكر ، عمان .
- محاسنة ،ابراهيم محمد (2013) : القياس النفسي في ظل النظرية التقليدية والنظرية الحديثة ،ط1 دار جرير للنشر والتوزيع عمان ،الاردن.
- عمر ،محمود وفخرو ، حصه والسبيعي ،تركي وتركي ،امنه(2010): القياس النفسي والتربوي ،ط1،عمان ،دار الميسرة للنشر والتوزيع .
- مراد، صلاح احمد وسليمان ،علي امين (2005) : الاختبارات والمقاييس في العلوم النفسية والتربوية خطوات اعدادها وخصائصها ،ط2،دار الكتاب الحديث، القاهرة .

- عزيز، صادق عبد النور (2015): اثر ترتيب فقرات مقياس الاتساق المعرفي على الخصائص القياسية عند الطلبة المرحلة الاعدادية وفقا لنظرية الاستجابة للفقرة ، اطروحة دكتورا غير منشوره ،جامعة بغداد ،كلية التربية ابن رشد .
- سوالمة، يوسف (1994): اختبارات الصواب والخطأ المتعدد الخصائص النفسية وعدد البدائل ،**المجلة الاردنية في العلوم التربوية** ، جامعة اليرموك الجزء (10) العدد(3) ص 573-595.
- مخائيل، امطانيوس .(1997). **إختبارات الذكاء والشخصية**، ج1، دمشق، منشورات جامعة دمشق.
- ابو علام، رجاء .(1989). **مدخل الى مناهج البحث التربوي**، الكويت، مكتبة الفلاح.

عبد الرحمن، سعد .(1983). **القياس النفسي**، الكويت، مكتبة الفلاح

- علام ،صلاح الدين : (2001).**الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية** ، القاهرة ، دار الفكر العربي
- عودة، احمد، مكاوي، حسن فتحي .(1987). **أساسيات البحث العلمي في التربية والعلوم الإنسانية**، عمان، مكتبة المنار.

المصادر الاجنبية:

Cronbach, L.J. and Gleser, G. (1970): Essentials of psychological - testing, 3rd ed., New York, Harper and row publisher.

-Ebel,R.L.&Frisbie,D.A(1991)Essentials of Educational Measurment(5th Ed.),Prentice-Hall Englewood Cliffs.

-Tsai,F.&Suen,H(1993)Abriefreport on a Comparion of Six Scoring methods for multiple true-false items, Educational and psychological Measurement,53(2),399-404.

- Crocker, L. & Algine, J. (1986) **Introduction to Classical and Modern Test Theory**. 2nd. New York. Holt, Rinehart & Winston.

- Anastasi, A. (1976) **Psychological Testing**, 4thed. New York, Macmillan

- Brennan, R . L . (1972) : **A** Kline, P.(1979) : **psychometric and psychology** ,London: Academic, press.

-Generalized Upper-Lower Item Discrimination Index. Educational and Psychological Measurement. Vol. 32. 209-303.

-Freeman. F.S. (1962): **Theory and Practice of Psychological testing**, 3rd ed, New York, Holt, Rinehart and Winston.

-Kline, P.(1979) : **psychometric and psychology** ,London: Academic, press.

- Magnusson, D (1967): **Testing theory**, London, Addison – Wesley Publishing Company.

- Rasheed, R. Y. R., & Sabri, D. A. S. (2022). Impact of Google workspace for Education platform on developing self-learning skills for applicators of Arabic language departments in colleges of education in Kurdistan region. Baltic Journal of Law & Politics, 15(2), 751-762