[١]

# اشتقاق معايير لتقدير الكفاءة النسبية للاختبار التحصيلي البنكي في ضوء بعض المؤشرات الإحصائية والمعرفية (باستخدام نظرية الاستجابة للمفردة)

**د. محمد منصور محمد الشافعي**

**أستاذ القياس والإحصاء المشارك بكلية التربية**

**جامعة الملك سعود**

# اشتقاق معايير لتقدير الكفاءة النسبية للاختبار التحصيلي البنكي في ضوء بعض المؤشرات الإحصائية والمعرفية
# (باستخدام نظرية الاستجابة للمفردة)

د. محمد منصور محمد الشافعي ★

**ملخص:**

استهدفت الدراسة اشتقاق معايير لتقييم الكفاءة النسبية للاختبار التحصيلي البنكي وهو الاختبار الذي سحبت مفرداته من بنك للأسئلة يحتوي على مفردات اختبارية يتوافر لها معلومات وخصائص إحصائية محددة بشكل مسبق، وقد حُددت تلك المعلومات والخصائص من خلال التحليلات الخاصة بأحد نماذج نظرية الاستجابة للمفردة التي استخدمت في إنشاء ذلك الأسئلة، وقد سعت الدراسة لاشتقاق هذه المعايير اعتمادا على مجموعة من المتغيرات الإحصائية والمعرفية التي تتعلق بمفردات الاختبار وهي: مستوى الصعوبة، قوة التمييز، مستوى التخمين، أقصى معلومات، تمثيل نطاق المتغير المقاس، ونوع المفردة، المستوى المعرفي، وقد اعتمدت الدراسة في إجراءاتها على إعداد بنك مصغر للأسئلة يتكون من (١٨٠) مفردة اختبارية في المهارات الإحصائية، بعد تحليل بياناته التي اشتقت من تطبيقه على عينة حجمها (٧٢٢) فرد من طلاب وطالبات مرحلتي البكالوريوس والماجيستير بكلية التربية بجامعة الملك سعود باستخدام التحليلات الخاصة بنموذج التقدير الجزئي (راش في

★ أستاذ القياس والإحصاء المشارك بكلية التربية جامعة الملك سعود.

صورته العامة) وايضاً التحليلات الخاصة بالنموذج الثلاثي البارمتر (وذلك للحصول بعض المؤشرات الإحصائية التي لاتتوافر بتحليلات نموذج التقدير الجزئي)، وقد قامت الدراسة بسحب عينة من مفردات البنك المصغر بلغ حجمها (١٤٤) مفردة لإعداد اختبار الدراسة في صورته النهائية وذلك في ضوء الضوابط الخاصة بمعالجة متغيراتها البحثية، وقد اعتمدت الدراسة في اشتقاقها لمعايير تقييم الكفاءة النسبية للاختبار التحصيلي البنكي على تحليل بيانات اختبار الدراسة البنكي باستخدام عدد من الأساليب الإحصائية وهي: التحليل التمييزي، معادلة نانلي لتقدير الدرجات الحقيقية، ومعامل إرتباط الرتب لسبيرمان، وقد اقترحت الدراسة في ضوء نتائج الإجراءات والتحليلات الخاصة بها درجات موزونة لكل مستوى محدد من مستويات كل متغير بحثي في ضوء مدى وحجم اسهامه في التمييز بين المستويات التحصيلية المختلفة لأفراد عينة الدراسة في ضوء المعدل التراكمي لهم.

وقد توصلت الدراسة في النهاية إلى اشتقاق معايير في ضوء هذه الدرجات الموزونة يمكن من خلالها تقدير الكفاءة النسبية للاختبار التحصيلي البنكي، كما يمكن استخدامها ايضا في تقدير كفاءة أي اختبار تحصيلي شريطة أن يتوافر لمفرداته معلومات وخصائص إحصائية وصفات معروفة بشكل مسبق.

## Abstract:

Study aimed to derive the criteria for evaluating the relative efficiency of the test grades banking a test that pulled its items from a item bank which has items with available information and statistical properties defined in advance, determined by information, and properties through the analysis of one of the theoretical models of (IRT) that Used in the establishment of such items, the study sought to derive these criteria based on a set of statistical and cognitive variables and related to test items: the level of difficulty, the power of discrimination, the level of guessing, the maximum information, the representation of measured variable domain, and type of item test, cognitive level, have been adopted study in their actions on the preparation Bank microcosm of the questions consists of (180) test items in statistical skills, after analyzing the data derived from its application to sample size (722) individual students undergraduate and master at faculty of Education, King Saud University using the analyzes of the Partial Credit Model (Rush in his public image) and also the analysis of the form Three Parameter Logistic Model (so as to get some statistical indicators that not available in analysis model estimate partial), The study to withdraw a sample of the items of the bank mini size was (144) Single to prepare a test study in its final form, in light controls for the treatment of research variables, the study has been adopted in the derivation of criteria for evaluating the relative efficiency of the test grades banking on the analysis of test data bank study using a number of statistical methods, namely: Analysis of Nanli equation to estimate the true scores, and coefficient of rank

correlation of Spearman, has proposed the study in light of the outcome measures and analyzes its own scores measured for each level of specific levels of each variable research in light of the extent and the size of his contribution to the distinction between the levels of realizable various members of the study sample in light of the cumulative average for them, The study found in the end to derive the criteria in the light of these grades weighted by which to estimate the efficiency of a relative of the test grades the bank, can also be used a well in estimating the efficiency of any achievement test, provided that available to vocabulary information and statistical properties Recipes are known in advance.

## المقدمة والخلفية النظرية:

تمثل الاختبارات التحصيلية التقويمية دوراً هاماً وبارزاً في مجال تقويم أداء المتعلمين، فهي الأدوات التي يتم من خلالها الكشف عما انجزه المتعلم من خبرات وما اكتسبه من مهارات وما حققه من أهداف في محيط المادة التعليمية المستهدفة من الاختبار (Hurwitz, Elliott, & Braden, 2007; Meisinger, Bradley, Schwanenflugel, Kuhn, & Morris, 2009)، كما أنها تعتبر من الركائز التي يعتمد عليها في إصدار العديد من القرارات في مختلف المؤسسات والهيئات التعليمية والأكاديمية (Hurwitz et al., 2007)، وقد أجريت العديد من الأبحاث والدراسات العلمية بهدف التوصل إلى الشروط والضوابط التي تجعل من هذه الأدوات وسائل موثوق بها، وعلى الرغم من تعدد وتنوع البحوث والدراسات المعنية بهذا اللأمر وما أفضت إليه من أطر تنظيرية تتعلق بطرق وشروط وضوابط الإجراءات الخاصة بإعداد الاختبارات التحصيلية (الزيد، ٢٠٠٧)، وما يتصل بها من متغيرات تتعلق بنوع مفردات الاختبار، ومستوى صعوبتها وقوتها التميزيية (Tasdemir& Mehmet, 2010) ودوال المعلومات الخاصة بها، Lihua Yao (2011) والمستويات المعرفية المستهدف قياسها Kucuk & Walters (2009)، ومدى تمثيل الاختبار للنطاق المستهدف من القياس، وطول الاختبار، Magis, David (2011) والزمن المناسب للإجابة على فقرات الاختبار، ومدى تحقق الشروط السكومترية للقياس (ياسين، ٢٠٠٤) Tavakol, Mohsen; Dennick, Reg (2011)،

وتوزيع الدرجات على فقرات الاختبار De la Torre, Jimmy; Yuan Hong, Hao Song (2011) وإعداد جدول المواصفات

(Aaron, Irina & David,2005) إلا أن العديد من الاختبارات مازالت تعاني حتى هذه اللحظة من بعض نواحي الضعف وأوجه القصور التي تشكل عوائق تحول دون تحقيق هذه الاختبارات لأهدافها ((De Champlain, 2010: 109)، ونظراً لأن عملية إعداد وبناء الاختبار التحصيلي التي ترمي إلى تحقيق القياس الموضوعي للجانب التحصيلي هي عملية على درجة عالية من الفنية تتطلب خطوات تجريبية تهدف إلى التحقق من الشروط الواجب توافرها في الاختبار الموضوعي (نصار، ٢٠١٠: ١٩٧). فقد سعت بعض الأوساط التعليمية إلى إنشاء بعض المؤسسات العلمية والأكاديمية المتخصصة في مجال بناء وإعداد الاختبارات التحصيلية مثل هيئة الاختبارات التربوية الأمريكية Educational Testing Service (ETS) المكتب الاسكتلندي لدائرة التربية Scottish Office Education Department (SOED) والمركز القومي للامتحانات والتقويم التربوي بمصر، National Center of Examination and Educational Evaluation (NCEEE) وهي المؤسسات المعنية بهذا المجال وتسعى إلى إعداد اختبارات تحصيلية على درجة عالية من الدقة.

وقد تبين للباحث الحالي من خلال خبراته المهنية التي اكتسبها من في مجال القياس ومن خلال اشتراكه كعضو في لجنة تقويم اختبارات الرياضيات بالثانوية العامة لعدة أعوام أن عملية تقييم هذه الاختبارات تقتصر في عرضها على بعض المختصين في مجال هذه الاختبارات، ونظراً لعدم توافر أية مؤشرات ومواصفات إحصائية مسبقة عن المفردات المتضمنة بتلك الاختبارات فإن عملية تقويم تلك الاختبارات تعتمد على مجرد الفحص الذاتي والاجتهاد الشخصي من جانب المقوم المتخصص

بمادة الاختبار، وقد يقتصر في الوقت الراهن على هذا الإجراء نظرا لافتقار العملية التقويمية لهذه الاختبارات إلى اسلوب ونظام تقييمي يستند إلى معايير وفقا لأسس علمية مدروسة يمكن من خلالها الحكم الموضوعي على كفاءة هذه الاختبارات في تحقيق الغرض منها، وكنتيجة لهذه المشكلة نشأت فكرة بنوك الأسئلة (عودة، ٢٠١٠، ٣٧٦) التي تحتوي أنواع مختلفة من المفردات التي يجرى تدريجها على متصل المتغير موضوع القياس Huang (2010: 475) كما أن إنشاء البنك يقتضي تمحيص وتدقيق كل مفردة بنكية من خلال إجراءات علمية دقيقة تسفر عن مؤشرات إحصائية ومواصفات تدون بالكرت البنكي الخاص بكل مفردة Item Card Shechtman, Lehman, Woodbury &ويستطيع واضع الاختبار من خلال الكروت البنكية التي تحتوي على المواصفات المتعلقة بالمفردات البنكية (مستوى الصعوبة، مستوى التمييز، مستوى التخمين،...) أن يختار من بينها في حبك وتفصيل الاختبار البنكي Tailored test، (كاظم، ١٩٨٨: ١٢٣-١٢٤)، Katie (2011)، والذي يتسم بمواصفات وخصائص محددة، وعلى الرغم من أن بنوك الأسئلة توفر الوقت والجهد لواضع الاختبار في سحب المفردات البنكية بالمواصفات التي يرغب بها، إلا أن هناك بعض المتغيرات الهامة التي ينبغي تمحيصها من الوجهة البحثية والتي يمكن من خلالها تحديد الكفاءة النسبية للاختبار البنكي Huang (2010: 376)، وبالتالي كان على واضع الاختبار البنكي أن يجب عن أسئلة تتعلق بهذه المتغيرات والتي قد تمثل الإجابة عنها معاييراً يمكن الاحتكام إليها عند تقييم هذه الاختبارات من حيث كفاءتها النسبية وهي: ما المستويات المناسبة من متغيرات الصعوبة والتمييز والتخمين وأقصى

معلومات؟ التي تحقق الكفاءة النسبية للاختبار التحصيلي البنكي؟، وما أنواع المفردات ومستواها المعرفي ومدى تمثيلها لنطاق المتغير المقاس التي تحقق أيضاً الكفاءة النسبية لنفس الاختبار؟

وبالتالي فإن الإجابة عن التساؤلات السابقة قد تلبي حاجة ملحة إلى اشتقاق معايير جديدة، قد تفيد في انتقاء مفردات الاختبار التحصيلي البنكي ومن ثم التقييم الكمي لكفاءته النسبية، وبالتالي جاءت فكرة البحث الحالي التي ترمي إلى الإجابة عن تلك الأسئلة وذلك من خلال اختبار القدرة التتنبؤية لكل مستوى من مستويات كل متغير من المتغيرات المتضمنة بالتساؤلات السابقة في التمييز بين مرتفعي ومنخفضي التحصيل، ودرجة اسهام كل متغير من المتغيرات المستهدفة بالدراسة في التمييز بين هاتين المجموعتين من الأفراد، وكذلك اختبار قوة ارتباط تصنيف وترتيب كل من الدرجات الحقيقية لأفراد عينة الدراسة وتقدير قدراتهم– في ضوء المعالجات الخاصة بكل متغير من متغيرات الدراسة– بتصنيف وترتيب نفس الأفراد في ضوء بيانات محكية تتمثل في معدلهم التراكمي، ومن ثم إمكانية التوصل إلى نموذج تقييمي كمي اعتمادا على أسس علمية مدروسة.

وبالتالي فإن هذا النموذج في حال التوصل إليه قد يلبي رغبة بعض المؤسسات والهيئات المختصة التي يوكل إليها أمر تقييم الاختبارات التحصيلية بصورة عامة والبنكية بصورة خاصة في استخدام معايير علمية قد تفيد في إجراءات تقييم الكفاءة النسبية هذه الاختبارات.

**الاختبار التحصيلي:**

الاختبار التحصيلي أحد المكونات الرئيسية للعملية التعليمية (ملحم، ٢٠١١: ٢٠٥)، ويقصد بالاختبار التحصيلي كما تعرفه أنستازي بأنه مقياس موضوعي مقنن لعينة من السلوك (Anastasi,1988:233)، كما يعرف الاختبار التحصيلي بأنه طريقة منظمة لتحديد مستوى تحصيل الطالب لمعلومات ومهارات في مادة دراسية كان قد تعلمها مسبقا بصفة رسمية، من خلال إجاباته على عينة من الأسئلة تمثل محتوى المادة الدراسية (نصار، ٢٠١٠: ١٩٨).

**الاختبار التحصيلي البنكي:**

وهو الاختبار التي تسحب مفرداته من رصيد المفردات البنكية المحتواه ببنك الأسئلة في ضوء جدول المواصفات وبعض المؤشرات الإحصائية والذي يمكن من خلالها حبك وتفصيل الاختبار البنكي في ضوء اهداف الاختبار وطبيعة المختبرين (كاظم، ١٩٨٨: ١٢٣-١٢٤).

**تقدير مواصفات ومؤشرات المفردات البنكية:**

نظرا لأن الدراسة الحالية تعتمد على بعض المؤشرات الإحصائية في اشتقاق معايير يمكن من خلالها تحديد الكفاءة النسبية للاختبارات التحصيلية التي تحتوي على انواع مختلفة من الأسئلة كان من الضروري انتقاء نماذج رياضية تناسب طبيعة الأنواع المختلفة من الأسئلة يمكن من خلالها تحديد بعض المؤشرات الإحصائية المعنية بهذه الدراسة وأحد هذين النموذج هو النموذج ثلاثي البارامتر Three Parameter Logistic Model التي تسفر تحليلاته للمفردات الاختبارية ثنائية التقسيم عن بعض المؤشرات التي يمكن الاعتماد عليها في المعالجات

الخاصة بالدراسة الحالية مثل مؤشرات الصعوبة، التمييز، التخمين، أقصى قيمة للمعلومات، وأيضا نموذج التقدير الجزئي Partial Credit والذي يناسب الأسئلة المقالية وهي الأسئلة متعددة الإجابة (التقسيم) والذي تسفر تحليلاته عن بعض المؤشرات الإحصائية مثل الصعوبة وأقصى معلومات وفيما يلي وصف مختصر لهذين النموذجين.

**النموذج ثلاثي البارامتر:**

**Three-Parameter Logistic Model:**

يفترض هذا النموذج أن بيانات الاختبار تتفاوت في صعوبتها وفي قوتها التمييزية وكذلك في احتمالية التخمين لإجاباتها (De Ayala, 2009: 123-130)، ويعتمد هذا النموذج على دالة رياضية ترتبط باحتمالية الاستجابة الصحيحة على المفردة، والتي تعني احتمالية الاستجابة الصحيحة على المفردة P(θ). وذلك على افتراض أن P(θ) تزيد بزيادة القدرة (θ). وتمثل الدالة في النموذج ثلاثي البارامتر بالصيغة التالية:

$$P(\theta) = c + \frac{1-c}{1+e^{-1.7a(\theta-b)}}$$

حيث ترمز a، b، c للبارامترات المميزة للمفردة كما يلي: فيمثل البارامتر a القوة التمييزية للمفردة Discriminating Power أما b فهي صعوبة المفردة Item Difficulty والبارامتر c يمثل الخط التقاربي الأسفل Lower Asymptote للمنحنى المميز للمفردة وهو بارامتر التخمين Guessing Parameter.

**نموذج التقدير الجزئى Partial Credit Model:**

ويصلح هذا النموذج لمفردات المقال التى تتضمن مستويات متدرجة ومتعددة من الأداء، وهذا النموذج له نفس ألاساس الرياضى الاحتمالى الخاص بجميع نماذج (IRT) اللوغاريتمية وبالتالى فإن الصورة العامة لهذا النموذج التى تعبر عن حصول الفرد (n) على الدرجة (x) للمفردة (i) هى (De Ayala,2009,163)،

$$\varphi_{nix} = \frac{\exp \sum_{j=0}^{x} (\theta_n - b_{ij})}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} (\theta_n - b_{ij})}$$

x = o,l..... m,

$$\sum_{k=o}^{mi} (\theta_n - bij) = 0 \qquad b_{ij} = 0$$

حيث

$$\exp \sum_{j=o}^{x} (\theta_n - b_{ij}) = 1$$

حيث $\phi_{nix}$ احتمال أن يحصل الفرد (n) على الاستجابة (x) عن المفردة (i) $b_{ij}$ هي صعوبة الخطوة j حيث(j=1,2,3.....mi) (المفردة i، حيث k هي عدد خطوات الحل حيث (k=1,2,3,.... mi).

والنموذج السابق يستخدم فى تحليلات برنامج Winsteps لأسئلة المقال، (Linacre, 2005: 9) وسوف يقتصر الباحث الحالى على استخدام نموذج بارشال كريدت وهو نموذج راش في صورته العامة، نظراً لأن هذا النموذج يناسب تحليل جميع أنواع مفردات الاختبار المستخدم بالدراسة الحالية وهي المفردات الموضوعية ثنائية التقسيم وأيضا المفردات المقالية متعددة التقسيم (Linacre, 2005: 15).

تأثير بعض المتغيرات الخاصة بالاختبار التحصيلي على أداء الممتحنين لقد بدا من واقع بعض الأطر النظرية والأدبيات بعض التاثيرات المحتملة لبعض المتغيرات على الأداء الاختباري وهي:

**مستوى صعوبة فقرات الاختبار:**

يرتبط مستوى صعوبة الفقرة الاختبارية بالمستوى العقلي المطلوب وبمستوى الهدف المعرفي، حيث تعتبر الفقرة الاختبارية المصنفة في أدنى المستويات المعرفية وهو أسهل الفقرات الاختبارية وتعتبر الفقرة الاختبارية المصنفة في المستويات العليا وفقا لتصنيف بلوم للأهداف من أصعب الفقرات الاختبارية نظرا لتعقد العمليات العقلية المرتبطة بكل مستوى من تلك المستويات (عودة، ٢٠١١: ٧٧)، وقد تبين من خلال التراث أن اغلب اسئلة الاختبارات تركز على المستوى المتدنى في تصنيف بلوم.

فقد أظهرت دراسة Aaron, Irina & David (2005) أن أغلب أسئلة الاختبارات التحصيلية تركز فقط على الجانب المعرفي الذي يتعلق بالمستويات الدنيا وفقا لتصنيف بلوم مثل التذكر والفهم والتطبيق، أما الأسئلة التي تتعلق بالمستويات العليا بتصنيف بلوم والتي تتعلق بالتحليل والتركيب والتقويم يندر الاهتمام بها في تلك الاختبارات على الرغم من ارتباطها بالتقويم الحقيقي لعملية التعلم والتي ينبغي أن يكون لها حظا من أسئلة تلك الاختبارات، وقد تشير المستويات العليا في تصنيف بلوم إلى كفاءة الاختبار.

حيث أظهرت دراسة Kucuk & Walters (2009) أن إعتماد الاختبار على المستويات المعرفية العليا وفقا لتصنيف بلوم هي مؤشرا

آخر على كفاءة الاختبار، وكذلك نوعية المفردات التي يعتمد عليها الاختبار والتي تبتعد إلى حد ما عن التخمين، كما اشارت ايضا إلى بعض المؤشرات الإحصائية التي يمكن الاستناد إليها في قياس كفاءة الاختبار مثل مؤشرات الصعوبة وبصفة تلك المؤشرات التي تتعلق بالمستويات المتوسطة الصعوبة وبعض المؤشرات التي تتعلق بالأسئلة الصعبة وونفس الحال مع مؤشر التمييز، كما حاولت بعض الدراسات الكشف عن علاقة نوع المفردات الاختبارية بمستوى الصعوبة، وقوة التمييز حيث أبرزت دراسة Tasdemir & Mehmet (2010) أن أسئلة الصواب والخطأ ليست اسهل من اسئلة الاختيار من متعدد، كما توصلت الدراسة أن إحتواء الاختبار على اسئلة متنوعة من الاختيار من متعدد والصواب والخطأ والمزاوجة له آثار إيجابية على ثبات الاختبار، كما انتهت الدراسة أن أسئلة الاختيار من متعدد والصواب والخطأ والمزاوجة لها نفس الخصائص الخاصة بالصعوبة والتمييز التي تتمتع بها أسئلة الاختيار من متعدد، وبالتالي فإن اختبار الاختيار من متعدد ليس له الأفضلية على اسئلة الصواب والخطأ أو المزاوجة في قياس المستوى التحصيلي للطلاب، وهذا ما أكدته دراسة (Jandaghi & Shateria, 2008).

كما استهدفت بعض الدراسات الكشف عن تأثر صعوبة أسئلة الاختبار بمستوى ثبات وصدق الاختبار حيث أظهرت بعض الدراسات مثل دراسة Jandaghi & Shateria (2008) عدم وجود فروق جوهرية بين صعوبة الأسئلة مع النسب العليا للثبات والصدق للاختبارات بينما وجدت أن نوعي أسئلة الصواب والخطأ وكذلك الاختيار من متعدد لهما نفس القوة التنبؤية الخاصة بصدق الاختبار، كما أن الدرجات

الحقيقية التي يمكن الحصول عليها من هذين النوعين من الاختبارات كانت متكافئة في حدود الخطا المعياري للقياس، وتعتبر أسئلة الصواب والخطأ بديل جيد لأسئلة الاختيار من متعدد، كما أن مستوى صعوب أسئلة الاختبار يعتبر من بين المتغيرات التي ترتبط بالدرجة الحقيقية التي يمكن أن يحصل عليها متلقي الاختبار حيث أوضحت دراسة Tavakol, et al (2011) بعض المؤشرات التي يعتمد عليها في التعرف كفاءة الاختبار في حصول الطلاب على الدرجات الحقيقية ومنها مستوى صعوبة بنود الاختبار ومستوى تمييز تلك البنود وكذلك المستوى المعرفي للأسئلة، وقد توصلت الدراسة إلى أن الأسئلة الموضوعية مؤشرا جيدا يمكن الاعتماد عليه في قياس قدرة الممتحن على التفسير، كما أن توظيف الطالب للمعلومات في معالجة بعض المواقف الجديدة مؤشرا جيدا في قياس ابتكارية الممتحن، كما أن تحديد مؤشرات الصعوبة للبنود تساهم في إعداد اختبار يمكن أن يستخدم بكفاءة في تقويم الأداء الأكاديمي.

**مستوى تمييز فقرات الاختبار:**

يرتبط مستوى تمييز الفقرات الاختبارية بمفهوم الفروق الفردية بين الأفراد، حيث يلعب مستوى تمييز الفقرة أدوار غاية الأهمية في تحديد الأفراد الذي حققوا وانجزوا الأهداف التعليمية بكافة مستوياتها، وامتلاك الفقرة الاختبارية للمستوى المتوسط من التمييز يؤدي إلى تصنيف الأفراد بمستويات عديدة من متغير القياس تنحصر بين المستويات الدنيا وكذلك العليا من نفس المتغير(Kingsbury & Zara, 2002: 230)، وعلى الرغم من أن الكثيرون يعتقدون بان المستوى المتوسط من التمييز هو

الأفضل عند انتقاء الفقرات الاختبارية إلا أن البعض يعتقد أن بعض الفقرات الاختبارية ينبغي أن تمتلك مستويات مرتفعة لتصنيف الأفراد بكافة مستويات المتغير المستهدف قياسه من الاختبار. Richard, (2001:214) وقد أجريت بعض الدراسات للتعرف على بعض انواع الأسئلة التي تتمتع بقدرة عالية على التمييز حيث استهدفت دراسة Schriner & Clifton (2010) التعرف على كفاءة أسئلة الاختيار من متعدد في قياس مستويات التعلم المعرفية المختلفة (من التذكر إلى التقويم)، وكذلك قدرة هذه النوعية من المفردات على قياس التفكير الناقد، وقوة التمييز الخاصة بهذه النوعية من الأسئلة، وقد اسفرت نتائج الدراسة إلى أن هذه الأسئلة لها القدرة على قياس التفكير الناقد، وقياس المستويات المعرفية العليا ولكنها ليست نفس القدرة التي يمكن أن تتمتع بها اسئلة المقال، وهذه النوعية من الأسئلة لها القدرة على التمييز بين المستويات التحصيلية المختلفة، وفي هذا الاطار أيضا أضافت دراسة (Tasdemir & Mehmet, 2010) أن أسئلة الصواب والخطأ لها نفس الخصائص الخاصة بالتمييز التي تتمتع بها أسئلة الاختيار من متعدد والمزاوجة، وقد أشارت بعض الأدبيات إلى قياس الاختبار لمهارات تفيد بكفاءته، حيث أشارت دراسة Ryan, Peter & (Elizabeth, 2010) إلى بعض المهارات يمكن من خلالها الحصول على دليل لصدق اختبار Idaho Alternate Assessment (IAA) حيث أظهرت نتائج الدراسة أن أكثر الأدلة صدقا في التمييز بين الطلاب هي الأدلة المرتبطة بالمهارات الأكاديمية ومهارات التفكير الناقد وكذلك المهارات الحسابية، هذا وترتبط قوة تمييز بنود الاختبار بالدرجة الحقيقية الذي يمكن الحصول عليها من الاختبار حيث أظهرت دراسة

Tavakol et al (2011) أن بعض المتغيرات تؤثر على الدرجة الحقيقية التي يمكن أن يحصل عليها الطالب في الاختبار مثل مستوى تمييز تلك البنود، كما ويتاثر مؤشر التمييز سلبياً ببعض المتغيرات حيث أشارت دراسة (Richard, 2001) إلى أن نسبة التخمين المرتفعة تؤثر سلبا في تقدير التمييز للمفردات، وكذلك تعددية البعد للاختبار تؤثر في تقدير قوة التمييز للمفردات بصورة سلبية.

**مستوى التخمين لفقرات الاختبار:**

يساهم متغير التخمين في حصول متلقي الاختبار على درجات إضافية قد تؤدي في النهاية إلى الإضرار بدقة تقديرات القدرة الخاصة بالمستجيب اعتمادا على تلك الفقرات، وذلك نظرا لأن الاستعانة بهذا الأسلوب في التعامل مع الفقرات الاختبارية يكون من قبل الممتحن الذي لايمتلك الخبرة او المعلومة أو الإمكانية التي تمكنه من الإجابة صواباً على تلك الفقرة (Burton, 2005).

وبالتالي فإن وجود فقرات اختبارية يتفاوت إمتلاكها لمستويات مختلفة من التخمين يمكن أن يؤدي في النهاية إلى حصول متلقي هذه الفقرات على درجات وتقديرات ليست دقيقة وبالتالي يمكن أن يؤدي هذا الأمر إلى الوقوع في اخطاء تتعلق بتصنيف هذا المتلقي بمستوى أو بفئة لاتمثل الحقيقة الواقعية والفعلية لمستواه، وبالتالي يمكن الاعتماد على ذلك المؤشر في تحديد مستوى كفاءة الاختبار في التمييز بين الفئات المختلفة من المتحنين، وقد كانت هناك محاولات من قبل الباحثين في تحديد بعض المؤشرات الدالة على سلوك التخمين، فقد اظهرت دراسة(Plake & Change Gene (2011 أن سلوك التخمين يمكن

تحديدة من خلال الإجابة الصواب عن عن اسئلة تمثل مستوى أعلى من القدرة المقاسة تفوق مستوى قدرة الممتحن، كما يتمثل بالجزء الذي يقطعه المنحنى المميز للمفردة Item Characteristic Curve (ICC) من المحور الرأسي الممثل لاحتمال حدوث الاستجابة الصواب على المفردة الاختبارية، كما أن هناك تأثيرات تحدثها الاستجابات الناتجة عن التخمين على بعض خصائص الاختبار السيكومترية والإحصائية، فقد اظهرت دراسة (Richard, 2001) إلى أن نسبة التخمين المرتفعة تؤثر سلبيا في تقدير التمييز للمفردات وكذلك ثبات الاختبار، وقد حاولت بعض الدراسات التعرف على التغير الذي يحدثه التخمين على اداء الطالب المتصل بانواع مختلفة من الفقرات الاختبارية، حيث أظهرت دراسة Lautenschlager, Shepler & Atwod Schurmeier (2010)، أن أداء الطلاب يتاثر بشكل دال بين صياغة الأسئلة المصاغة باسلوب الصواب والخطا عن أسئلة الاختيار من متعدد، وإذا صيغت بصورة تكملة أو مزاوجة، أوصيغة المقال عن الأنواع الأخرى من الأسئلة وقد أشارت الدراسة أن اسئلة المقال أفضل انواع الأسئلة في تقويم الأداء الحقيقي للمختبر، وقد فسرت الدراسة أن هذا الاختلاف قد يعزى إلى فرصة تخمين الإجابة التي قد ترتفع مع أنواع معينة من الأسئلة عن انواع أخرى من الأسئلة

**المستوى المعرفي لفقرات الاختبار:**

حدد "بلوم "Bloom تصنيفاً للأهداف التعليمة في المجال المعرفي، تبدأ من مستوى التذكر ثم الفهم فالتطبيق فالتحليل فالتركيب وينتهي بمستوى التقويم وهو يمثل المستوي المعرفي الأعلى في قمة

تصنيفه الهرمي للأهداف، وتلك المستويات المعرفية تؤخذ بعين الاعتبار عند بناء الاختبارات التحصيلية وذلك من خلال جدول مواصفات الاختبار (ملحم، ٢٠١١: ٢١٦)، وتتعلق إجراءات التقويم الحقيقي بمستويات معرفية محددة في تصنيف بلوم للأهداف حيث أظهرت دراسة (Aaron, Irina & David, 2005) بعض المستويات المعرفية والتي تستخدم في عمليات التقويم الحقيقي حيث توصلت الدراسة إلى أن البنود الاختبارية التي تقيس المستويات العليا هي التي تتعلق بمستويات التحليل والتركيب والتقويم بتصنيف بلوم وهي التي ترتبط بالتقويم الحقيقي لعملية التعلم والتي ينبغي أن يكون لها حظا من أسئلة تلك الاختبارات، كما اهتمت دراسات اخرى بالكشف على أنواع من الأسئلة التي تقيس مستويات مختلفة بتصنيف بلوم للأهداف في المجال المعرفي، فقد أظهرت دراسة Schriner & Clifton (2010) أن أسئلة الاختيار من متعدد لها القدرة في قياس المستويات المعرفية العليا ولكنها ليست نفس القدرة التي يمكن أن تتمتع بها أسئلة المقال، كما توصلت دراسة Tavakol, et al (2011) إلى أن توظيف الطالب للمعلومات في معالجة بعض المواقف الجديدة مؤشرا جيدا في قياس ابتكارية الممتحن وهي تلك المهارة التي ترتبط بالمستويات بالمستويات العليا في تصنيف بلوم للأهداف تساهم في إعداد اختبار يمكن أن يستخدم بكفاءة في تقويم الأداء الأكاديمي، وقد كانت هناك محاولات من جانب بعض الباحثين للكشف عن المهارات التي ترتبط بالمستويات المعرفية المختلف لتصنيف بلوم للأهداف حيث أظهرت دراسة Nilay (2007) أن مهارات المعرفة الواقعية Factual Knowledge ترتبط بالأسئلة التي تندرج تحت كل من مستويي التذكر والفهم، وأن المعرفة المفاهيمية Knowledge

Conceptual ترتبط بالأسئلة التي تصنف تحت مستوى الفهم وأن المعرفة الإجرائية Knowledge Procedural ترتبط بالأسئلة التي تندرج تحت مستوى التطبيق، وان عمليات مارواء المعرفة ومهارات التفكير الناقد Knowledge Meta-Cognitive ترتبط بالأسئلة التي تندرج تحت مستويي التركيب والتقويم. كما أجريت بعض الدراسات كشفت عن تقييم كفاءة الاختبار في ضوء بعض المستويات المعرفية المتضمنة بتصنيف بلوم للأهداف حيث أشارت دراسة Kucuk & Walters (2009) إلى أن المستويات المعرفية العليا وفقا لتصنيف بلوم هي مؤشرا جيدا على كفاءة الاختبار، وفي نفس الإطار أظهرت دراسة James (2011) أن عدد الأفراد الذين أنجزوا الأهداف التعليمية المرتبطة بالمقرر في ضوء اختبارات تحصيلي تم إعدادها في ضوء تصنيف بلوم تفوق على عدد الأفراد الذين انجزوا نفس الأهداف من المجموعة الأخرى التي تلقت اختبارات تحصيلية لم يتم إعدادها في ضوء تصنيف بلوم، كما أن توزيع البيانات الخاصة للدرجات التي تم الحصول عليها من الاختبارات التحصيلية التي أجري إعدادها في ضوء تصنيف بلوم للأهداف كانت مطابقة للتوزيع الطبيعي، وذلك فيما يتعلق بالاختبارات التي تتضمن أسئلة تندرج تحت المستويات التي تبدأ من مستوى التطبيق، كما أجريت بعض الدراسات للتعرف على ارتباط أسئلة الاختبار التحصيلي الجامعي بمستويات بلوم للأهداف، حيث اظهرت دراسة Bowers & Alex (2011) أن أغلب الاختبارات التحصيلية الجامعية اهتمت بالنواحي المعرفية الأكاديمية على حساب بعض النواحي الأخرى التي يمكن تكشف عن موهبة لدى الطالب، كما أنها لا

تتعلق بقدرة المتعلم على توظيف ماتم تعلمه بمواقف جديدة لم يسبق له أن مر بها من قبل وترتبط بالمستويات العليا من تصنيف بلوم للأهداف.

**أقصى قيمة لمعلومات فقرات الاختبار:**

**Maximum Information:**

وهو المؤشر الأساسي لدقة القياس في نظرية الاستجابة للمفردة (Edward, 2010: 378) وقد اشتق هذا المؤشر كدالة لمستوى القدرة المقدر للمفحوص وبارامترات المفردات المطبقة.(Lihua Yao, 2011) (Kingsbury & Zara, 2002). كما أن هذا المؤشر يدل على أقصى قيمة للمعلومات التي يمكن أن تعطيها المفردة الاختبارية والتي تتعلق بالأفراد عن مستوى معين من مستويات القدرة المقاسة، ( Rose, Bjorner, Becker & Friesc, 2008: 20) وبالتالي فإن المفردة الاختبارية التي تمتلك أعلى معلومات عند مستوى معين من القدرة المقاسة تكون أفضل من مفردات أخرى تمتلك معلومات أقل ( Edward 2010: 474).

وقد أثبتت العديد من الدراسات أن المفردات الاختبارية التي تمتلك المستوى الأعلى من متغير اقصى معلومات تعطي تقديرات ادق للقدرة المقاسة Saim Aslam, Muhammad, Atet (2011) مما يعني أن مؤشر أقصى معلومات يمكن أن يكون من بين المؤشرات التي يمكن الاعتماد عليها في تحديد الكفاءة النسبية للاختبار، وقد أشارت الأدبيات إلى تمتع اختبارات الموائمة المحوسبة بالعديد من المزايا من أهمها إمكانية تقدير قدرة المفحوص باستخدام عدد قليل من المفردات التي تتميز بامتلاكها لأقصى معلومات

(Siang& Fritz, 2006: 241)، كما أن اختيار المفردات بطريقة أقصى معلومات يوفر تقديراً للقدرة بأقل عدد من المفردات Schmitt, (2010: 233)، ولذلك يؤكد (Marinagi, & Kaburlasos, 2010) على أن أسلوب المعلومات القصوى يعطي نتائج جيدة في اختيار مفردات الاختبار ويؤدي إلى تحقيق مزيد من الدقة في قياس القدرة المستهدفة بالاختبار إذا قورن بأساليب أخرى في اختيار مفردات الاختبار.

كما أظهرت دراسة Saim Aslam, Muhammad, Atet (2011) أن أسلوب المعلومات القصوى في اختيار المفردات الاختبارية قد تفوق على أسلوب أقصى تمايز، ويشير Kingsbury & Zara, (2002) إلى عدة مميزات يتمتع بها أسلوب أقصى معلومات في انتقاء المفردات منها أنه يوفر أعلى كفاءة للاختبار، وكذلك تقدير القدرة يصبح أدق بصورة متزايدة مع تطبيق كل مفردة، كما أن انتقاء المفردة بأسلوب المعلومات القصوى كما يشير Lihua Yao (2011) يعطي أقل خطأ قياس عند مستوى القدرة المستهدفة، وتقدر قيمة أقصى معلومات يمكن أن تقدمها المفردة عن الأفراد عند مستويات محددة من القدرة وفقا لنموذج التحليل المستخدم والذي يتناسب مع نوعية معينة من المفردات، ونظرا لأن الدراسة الحالية سوف تعتمد على النموذجين (الثلاثي البارمتر) لتحليل البيانات الخاصة بالمفردات الموضوعية (ثنائية التقسيم) ونموذج التقدير الجزئي في تحليل البيانات المتعلقة بالأسئلة المقالية (متعددة التقسيم)، فإن الجزء التالي يعرض صيغ تقدير دوال المعلومات بهذين النموذجين:

**دالة المعلومات للنموذج الثلاثي البارمتر:**

وتقدر دالة المعلومات بنود الاختبار في هذا النموذج بالصيغة التالية (Linden& Hambleton, 2000) (Edward H. 2010):

$$I_i(\theta) = a^2 \left[\frac{Q_i(\theta)}{P_i(\theta)}\right]\left[\frac{P_i(\theta) - c^2}{(1-c^2)}\right]$$

حيث (Pi (θ هي احتمال حدوث الاستجابة الصواب للفرد الذي قدرته (θ)) على المفردة (i)، C تقديرا متغير التخمين على المفردة، a تقديرا لمتغير تمييز المفردة، Qi هي احتمال حدوث الاستجابة غير الصحيحة حيث: (Qi = 1.0- Pi (θ).

**دالة المعلومات في نموذج التقدير الجزئى:**

تقدر معلومات المفردة في نموذج التقدير الجزئى بالصيغة التالية Tie Liang W., Craig S. (2009):

$$I_i(\Theta) = \sum_{xi=0}^{mi} \frac{[P'_{xi}(\Theta)]^2}{P_{xi}(\Theta)}$$

حيث) Ii (θ هي دالة المعلومات للمفردة i عند مستوى القدرة $\theta$ ويوضح Lihua Yao & Schwarz (2006) أن المعلومات الكلية التي يقدمها الاختبار هي نفسها مجموع قيم معلومات كل مفردة.

**نوع الفقرات الاختبارية:**

تشتمل اختبارات التحصيل على أنواع مختلفة من المفردات فمنها ما هو موضوعي ومنها ماهو مقالي، وقد أشارت الأدبيات إلى شيوع استخدام الأسئلة الموضوعية في الاختبارات التحصيلية بالتعليم الجامعي

وبصفة خاصة نوعي أسئلة الاختيار من متعدد والصواب والخطأ حيث أوضح Burton (2005) أن غالبية الجامعات الأمريكية تعتمد بصورة أساسية على هذين النوعين، كما أشارت تلك الأدبيات إلى تأثر أداء الممتحن وكذلك ثبات الاختبار بنوع الأسئلة المتضمنه به فقد أظهرت دراسة مصطفى (٢٠٠٠) وجود فروق ذات دلالة إحصائية بين الدرجات الحقيقية لصالح أسئلة الصواب والخطأ إذا ماقورنت بأسئلة الاختيار من متعدد، وكذلك وجود فروق ذات دلالة إحصائية للدرجة الحقيقية للطالب بين مستويات الصعوبة (سهل- متوسط- صعب) في صالح المستوى المتوسط لأسئلة الاختيار من متعدد وكذلك مع أسئلة الصواب والخطأ، كما وجدت فروق جوهرية فى القيمة العددية لمعامل الثبات بين كل من أسئلة الاختيار من متعدد والصواب والخطأ لصالح أسئلة الاختيار من متعدد. كما أشارت بعض الأدبيات إلى تأثر بعض الخصائص الإحصائية للاختبار بنوع الأسئلة، فقد أوضحت دراسة Richard & Burton (2001) أن الاختبار الذي يقيس صفة متسعة المدى ينبغي أن يتم اختيار مفراته بناء على مؤشر التمييز، كما أن نسبة التخمين المرتفعة التي تتعلق بالأسئلة التي تحتوي على عدد أقل من البدائل تؤثر في تقدير التمييز للمفردات، وفي هذا الصدد أظهرت دراسة Richard (2004) أن ثبات الاختبار يتاثر بنوع الأسئلة، فقد تبين أن اسئلة الاختيار من متعدد لها تأثيرات إيجابية على ثبات الاختبار، وهو نفس الحال مع أسئلة المزاوجة، كما أن الثبات يزداد طرديا مع زيادة عدد خيارات أو بدائل الإجابة، كما أظهرت دراسة Jandaghi & Shateria (2008) أن أسئلة الاختيار من متعدد والصواب والخطأ لهما نفس القوة التنبؤية الخاصة بصدق الاختبار، وأن اسئلة الاختيار من متعدد تؤدي

إلى قيم أفضل نسبياً للثبات، وقد اظهرت النتائج أن الدرجات الحقيقية التي يمكن الحصول عليها من هذين النوعين من الأسئلة كانت متكافئة في حدود الخطا المعياري للقياس، كما أن اسئلة الصواب والخطأ تعتبر بديل جيد لأسئلة الاختيار من متعدد، وقد أظهرت دراسة Kuccuk & Walters (2009) أن نوعية المفردات التي يعتمد عليها الاختبار والتي تبتعد إلى حد ما عن التخمين يمكن أن تشكل مؤشرا جيداً لكفاءة الاختبار، كما أظهرت دراسة Schurmeier, et al (2010)، أن اداء الطلاب يتاثر بشكل دال بين صياغة الأسئلة المصاغة باسلوب الصواب والخطا عن أسئلة الاختيار من متعدد، وإذا صيغت بصورة تكملة أو مزاوجة، وكذلك إذا استخدمت صيغت بأسلوب المقال عن الأنواع الأخرى من الأسئلة كما تبين أن أسئلة المقال أفضل الأسئلة في تقويم الاداء الحقيقي، كما اظهرت دراسة Tasdemir & Mehmet (2010) أن أسئلة الصواب والخطأ والمزاوجة ليست أسهل من أسئلة الاختيار من متعدد، كما أن إحتواء الاختبار على أسئلة متنوعة من الاختيار من متعدد والمزاوجة والصواب والخطأ له آثار إيجابية على ثبات الاختبار، وانتهت الدراسة إلى أن أسئلة الاختيار الصواب والخطأ والمزاوجة لها نفس الخصائص الخاصة بالصعوبة والتمييز التي تتمتع بها أسئلة الاختيار من متعدد، وبالتالي فإن أسئلة الاختيار من متعدد ليس لها الأفضلية على اسئلة الصواب والخطأ أو المزاوجة، هذا وقد أشارت بعض الأدبيات بعلاقة أنواع أسئلة الاختبارات التحصيلية بأنواع التفكير وقياس المستويات المعرفية العليا وعمق المعرفة، حيث أظهرت دراسة Clifton (2010) & Schriner كفاءة أسئلة الاختيار من متعدد في قياس التفكير الناقد، كما أن لها القدرة في قياس المستويات المعرفية العليا ولكنها ليست نفس

القدرة التي يمكن أن تتمتع بها اسئلة المقال، ولها القدرة على التمييز بين المستويات التحصيلية المختلفة.

كما أظهرت دراسة Lewis, Munn, Jordon, Hicks, &Charles (2010) أن الاختبارات (المقالية) والتي تهتم بالنواحي التطبيقية بمقدورها إبراز النواحي الابتكارية، وهي أفضل الاختبارات في الكشف عن الخبرات التعليمية التي اكتسبها الطلاب في بالبرامج التعليمية، كما أظهرت دراسة Norbert & Janice (2011) فعالية استخدام طرق مختلفة لقياس لعمق المعرفة اللغوية باستخدام انواع مختلفة من الأسئلة، حيث أن هناك أنواع من الأسئلة التي يمكن أن تعكس عمق المعرفة اللغوية في هذه النوعية من الاختبارات مثل الأسئلة المقالية الانشائية، وكذلك أسئلة الاختيار من متعدد التي يمكن أن تحدد معان مختلفة لصيغ الكلمات هي أفضل الأسئلة التي يمكن من خلالها قياس المعرفة المتعمقة الخاصة باللغة.

**تمثيل المتغير المقاس:**

والمقصود من تمثيل المتغير المقاس هو تمثيل أسئلة الاختبار لأغلب الأهداف التعليمية المرتبطة بنفس المقرر، وغالبا ما يعتمد إعداد هذه النوعية من الاختبارات التي تمثل المقرر الدراسي على جدول المواصفات، لضمان تمثيلها غالبية موضوعات وأجزاء المقرر في ضوء مستويات الأهداف المرتبطة بتصنيف بلوم في المجال المعرفي Hudson,et. al., (2011)، وقد أجريت بعض الدراسات النادرة التي تناولت العلاقة بين تمثيل المقرر للنطاق السلوكي المستهدف من الاختبار ودقة نتائج القياس فقد اظهرت دراسة Swiatek & Mary

(2005) أن الطلاب المتفوقون تفوق أدائهم على الاختبارات الممثلة لنطاق القياس المستهدف بشكل جيد عن أداء الطلاب العاديون، بينما لم يختلف جوهريا عن اداء العاديون في الاختبارات التي لا تمثل النطاق السلوكي المستهدف بشكل جيد.

كما أظهرت دراسة Hudson, Stichter (2011)Lewis & أن بعض المؤشرات التي يمكن أن تميز جودة الاختبار من بينها تغطية أسئلة الاختبار لجميع موضوعات وفقرات المقرر الدراسي، وقد أشارت الدراسة أن أسئلة الاختبار ينبغي تعكس الأهداف واكثر موضوعية التعليمية المرتبطة بنسبة يجب أن تتجاوز النسبة (٦٠%) من جملة الأهداف التعليمة، كما أظهرت دراسة Tyrone & David (2011) أن الاختبارات المقننة التي تعدها الهيئات المتخصصة مثل مكتب خدمات الاختبارات التربوية الأمريكي Educational Testing (ETS) Services هي الاختبارات التي يمكن أن تقيس التعلم بصورة أدق نظراً لأنها اختبارات مقننة تحتوي مفردات تمثل السلوك المتعلم بصورة ادق وأشمل، كما أن هذه الاختبارات تحتوي على اسئلة من اختبارات تحصيلية تم إعدادها محليا واستخدمت على نحو سابق في الامتحانات الخاصة بتلك المؤسسات التعليمية كما أن البيانات التي استخدمت في تمحيص هذه الأسئلة هي بيانات حقيقية تم الحصول عليها من خلال تلك المؤسسات ومن واقع نتائج الامتحانات التي استخدمت بها تلك الأسئلة كما أن هذه الاختبارات مقننة وتتمع بدرجات مقبولة من الثبات والصدق، أما الاختبارات التحصيلية التي يتم انتاجها محليا هي اخبتارات لاتتمتع بخاصية تمثيل المجال المستهدف من القياس بنفس الدرجة التي تتمع بها الاختبارات المقننة التي تنتجها الهيئات المتخصصة، كما أن

التحقق من جودة هذه الاختبارات المحلية لا يكون خلال تقنيانات أو خصائص أو مؤشرات إحصائية يكون من يمكن من خلالها الثقة بتائجها ولكن يتم الاكتفاء بالتحقق من خصائص هذه الاختبارات من قبل فريق من المتخصصين والخبراء في مجال القياس المستهدف من تلك الاختبارات.

وبتضح مما سبق عرضه من دراسات أن تمثيل الاختبار التحصيلي لكل جوانب المقرر التحصيلي يعتبر أحد المؤشرات التي يمكن الاستناد إليها في الحكم على كفاءة الاختبار التحصيلي، كما أن نسبة الأهداف التي ينبغي أن تقاس من خلال أسئلة الاختبار ينبغي أن تتخطى النسبة (٦٠%) من جملة الأهداف التعليمة للمقرر المرتبط بالاختبار كما اوضحت ذلك دراسة Hudson,et. al. (2011) وهذا مما حدا بالباحث إلى معالجة متغير تمثيل الأسئلة لموضوعات المقرر بهدف الكشف نسبة تمثيل أسئلة الاختبار لموضوعات المقرر التي يمكن أن يعتمد عليها في تقييم الكفاءة النسبية للاختبار بالدراسة الحالية.

**مشكلة الدراسة:**

نتيجة لممارسة الباحث لعمله الأكاديمي والمهني بمجال القياس والتقويم وخبراته في مجال تقويم الاختبارات وبصفة خاصة الاختبارات التحصيلية، ونتيجة تفقده للأطر النظرية والأدبيات التي تتعلق بالقياس والتقويم، فقد تبين افتقار عملية تقييم الكفاءة النسبية للاختبارات التحصيلية في تحقيق ما ترمي إليه من أهداف إلى أسلوب علمي كمي دقيق، ونظر لأن بعض الاتجاهات الحديثة للقياس تعتمد على بعض النماذج الرياضية في تحليل مفردات مختلف أنواع الاختبارات وتسفر

عن مؤشرات إحصائية دقيقة لتلك المفردات، وحيث أن بعض الأدبيات أظهرت تأثير تلك المؤشرات الإحصائية للمفردات الاختبارية على أداء المستجيبين لهذه المفردات وكذلك تأثيرها على التقديرات الخاصة بمستويات قدراتهم والدرجات الحقيقية لهم، ومن بين هذه المؤشرات، مؤشر الصعوبة (2010) Tasdemir, Mehmet، Tavakol, et.al (2011)، مؤشر التمييز (2010) Ryan, Peter, Elizabeth,، مؤشر التخمين Schurmeier, et ، & Schriner Clifton (2010) دالة المعلومات Plake & Gene, (2011) Chang, ،al (2010) (Schmitt,2010:233)، Aslam & Altaf, Saim , Muhammad(2011)، فقد رأى الباحث الاعتماد على هذه المؤشرات في اشتقاق معايير إحصائية يمكن من خلالها الحكم على الكفاءة النسبية الخاصة بهذه الاختبارات في تحقيق الأهداف التي وضعت من أجلها، كما أن هناك بعض المؤشرات الأخرى التي يمكن الاعتماد عليها أيضا في تحديد الكفاءة النسبية للاختبار وذلك استنادا على بعض الأطر النظرية ذات الصلة، والتي أظهرت علاقة بعض المتغيرات الأخرى بكفاءة هذه الاختبارات ومن بين هذه المتغيرات متغير نوع المفردات، Tasdemir & Mehmet (2010) Kucuk & Walters(2009) Schurmeier, et al (2010) متغير المستويات المعرفية للأسئلة الاختبار Tavakol, et.al., ، & Schriner Clifton (2010) (2011)، متغير تمثيل الأسئلة للنطاق السلوكي Hudson, Lewis, Hudson, et. al., (2011) & Stichter (2011) والتي يمكن أن تؤثر بدورها على كفاءة هذه الاختبارات في تحقق الأغراض الخاصة بها، حيث تبدو تلك الكفاءة في قدرة الاختبار على التمييز بين

المستويات المختلفة للأفراد في ضوء المتغير المستهدف من القياس، وبالتالي يمكن التعبير عن هذه المشكلة في الإجابة عن السؤال العام التالي:

ما دلالات القوة التنبؤية لبعض المؤشرات الإحصائية لمفردات الاختبار التحصيلي (دالة المعلومات، الصعوبة، التمييز، دالة المعلومات، التخمين) وكذلك المستوى المعرفي للمفردات (من مستوى التذكر حتى المستويات المعرفية العليا) وايضاً نوع المفردة (اختيار من متعدد، صواب وخطأ، مزاوجة، إكمال، إجابة قصيرة، مقال) في التمييز بين المستويات التحصيلية المختلفة للمحنين.

**ويتفرع من هذا السؤال العام الأسئلة الفرعية التالية:**

١- ما دلالة القوة التمييزية لكل متغير من متغيرات (الصعوبة، التمييز، أقصى معلومات، التخمين) بين مرتفعي ومنخفضي المعدل التراكمي التحصيلي؟

٢- ما دلالة القوة التمييزية لكل متغير من متغيرات (نوع المفردات، تمثيل المفردات نطاق السلوك المقاس، المستوى المعرفي للمفردات) بين مرتفعي ومنخفضي المعدل التراكمي التحصيلي؟

٣- هل يختلف ترتيب أفراد عينة الدراسة وفقا لمعدلهم التراكمي مع كل من ترتيب تقديرات قدراتهم المشتقة من استجاباتهم على أسئلة الاختبار وترتيب درجاتهم الحقيقية على نفس أسئلة الاختبار التي صنفت في ضوء متغيرات الدراسة (الصعوبة- التمييز- دالة المعلومات- التخمين- نوع المفردات- نسبة تمثيل نطاق المتغير المقاس- المستوى المعرفي) بمستوياتها المختلفة.

٤- هل يمكن اشتقاق معايير كمية يمكن استخدامها في تقييم الكفاءة النسبية للاختبارات التحصيلية البنكية في ضوء متغيرات الدراسة؟

**أهمية الدراسة:**

- عدم وجود نموذج رياضي (في حدود علم الباحث) يمكن الاستناد إليه في تحديد الكفاءة النسبية للاختبارات التحصيلية إعتمادا على المؤشرات الإحصائية التي تتعلقب بكل من المفردات والأفراد.
- إن التوصل إلى نموذج رياضي إحصائي من شأنه توحيد المعايير والأسس العلمية التي يمكن استخدامها في إعداد الاختبارات التحصيلية مما يؤدي في النهاية إلى تحقيق الموضوعية المنشودة في مجال إعداد واستخدام الاختبارات التحصيلية.
- يمكن أن تسفر نتائج الدراسة الحالية عن نموذج رياضي يمكن من خلاله تقويم الاختبارات التحصيلية بصورة اكثر دقة وبناء على معايير يمكن أشتقاقها وفقا لأسس علمية مدروسة.

**أهداف الدراسة:**

- الكشف عن دلالة القوة التمييزية لبعض المؤشرات الإحصائية للاختبار المتمثلة في مستوى الصعوبة، قوة التمييز، ودالة المعلومات، والمستويات المعرفية، ومدى تمثيل المفردات للنطاق السلوكي، وطول وزمن الاختبار بين المستويات التحصيلية المختلفة للمتحنين.
- الكشف عن إمكانية اشتقاق صيغة رياضية لتحديد وتقدير الكفاءة النسبية للاختبارات التحصيلية في ضوء النتائج المتوقعة من الدراسة.

**مصطلحات الدراسة:**

*** الكفاءة النسبية للاختبار Relative test efficiency:**

ويحددها الباحث في قدرة الاختبار من خلال نواتج قياساته تقدير قدرات المختبرين الحقيقية والفعلية المستهدف قياسها من الاختبار منسوبة إلى الهدف من استخدام الاختبار.

فالاختبار التحصيلي الذي يستهدف الكشف عن المتفوقين له اهداف تختلف عن اهداف الاختبارات التحصيلية الأخرى التي تستخدم في تقويم التحصيل عن الطلاب بمختلف مستوياتهم، كما يختلف عن الاختبارات التحصيلية التشخصية التي تستهدف تحديد مشكلات التحصيل لدى الطلاب الضعاف تحصيلياً، وتعرف إجرائياً بقدرة الاختبار التحصيلي البنكي في التمييز بين فئتين أو أكثر من المختبرين المختلفين من حيث معدلهم التراكمي.

*** مستوى الصعوبة Difficulty level:**

ويتحدد بمواقع المفردات على متصل وتدريج الصفة المقاسة من الاختبار وتحدد امتلاكها لمستويات محددة من هذه الصفة (Jason, Edward & Walter, 2009). ويعرف إجرائيا بتقديرات الصعوبة المشتقة من تحليلات نموذج الاستجابة للمفردة اللوغاريتمي ثلاثي المعلم.

*** مستوى التمييز Discrimination level:**

هو قدرة المفردة على التمييز أو التفريق بين مستويات القدرة المقاسة لدى الممتحنين (Azevedo,2009) ويقدر في التحليلات الخاصة

بنماذج التحليل المرتبطة بنظرية الاستجابة للمفردة بظل الزاوية التي يصنعها المماس عند نقطة منتصف المنحنى المميز للمفردة على محور القدرة (Wang &Vispoel,1998: 116). ويعرف إجرائيا بتقديرات التمييز المشتقة من تحليلات نموذج الاستجابة للمفردة اللوغاريتمي ثلاثي المعلم

*** مستوى التخمين Guessing level:**

وهي المؤشرات التي تشير إلى توصل الممتحن إلى الإجابة الصحيحة عن المفردات التي لم تمكنه قدرته في الإجابة عنها عن طريق التخمين العشوائي Azevedo, (2009) Urry, (1977) Ponsoda, et al. (1997: 213). ويعرف إجرائيا بتقديرات التخمين المشتقة من تحليلات نموذج الاستجابة للمفردة اللوغاريتمي ثلاثي المعلم.

*** تمثيل نطاق القياس:**

**Measurement domain represented:**

متعرف بأنها نسبة تغطية مفردات الاختبار للأهداف التعليمية المرتبطة بموضوعات مقرر محدد (Hudson, et. al., 2011: 117)، ويعرف إجرائياً بنسبة تمثيل أسئلة الاختبار لمهارات الإحصاء كما وردت باختبار الدراسة الأصلي على اعتبار أن الاختبار الأصلي يمثل نطاق المتغير المقاس بنسبة (١٠٠%).

*** المستوى المعرفي Cognitive level:**

هو أحد المستويات المعرفية المستهدف قياسها من اسئلة الاختبار والمتضمنة بتصنيف بلوم للأهداف وتتمثل في مستويات (التذكر، الفهم،

التطبيق، التحليل، التركيب، التقويم) (James, 2011: 55)، ويعرف إجرائياً بتصنيف المحكمين لأسئلة الاختبار الأصلي تبعا لمستويات بلوم للأهداف في المجال المعرفي.

*** أقصى قيمة لمعلومات المفردات Maximum information:**

عرف فيشر Fisher المعلومات الإحصائية بأنها مقلوب الخطأ المعياري (De Ayala, 2009, 29)، في نظرية الاستجابة للمفردة فإنه من الممكن تحديد مقدار المعلومات التي يمكن أن توفرها المفردة عند كل نقطة خلال متصل (θ)، وكلما كان هناك معلومات أكثر (أعلى) عند مستوى محدد (θ)كلما كان هناك مقياس أدق عند نفس المستوى، وبالتالي فإن أحد مميزات نظرية الاستجابة للمفردة هي القدرة على تحديد أي المفردات تقيس بصورة أفضل عند مستوى (θ)بعينه Edward, (2010). ويعرف إجرائيا بتقديرات متوسط معلومات المفردة الاختبارية المشتقة من تحليلات كل من نموذجي التحليل (ثلاثي المعلم، التقدير الجزئي).

*** الدرجة الحقيقية True score:**

الدرجة يسفر عنها تطبيق الاختبار هي الدرجة الخام (الملاحظة) وتعتبر هذه الدرجة خليط من الدرجة الحقيقية بالإضافة إلى درجة خاطئة تسفر عنها أخطاء القياس وبطرح الدرجة الناتجة عن أخطاء القياس من الدرجة الخام (الملاحظة) يمكن الحصول على الدرجة الحقيقية، والدرجة الحقيقية يمكن الحصول عليها بإيجاد متوسط درجات الفرد التي حصل

عليها بموجب تطبيق الاختبار عليه لعدد كبير جدا من المرات، أو بتطبيق اختبار عليه كامل الدقة (Paul, 1991: 630).

ويعرف إجرائياً بالدرجات الحقيقية المستخرجة من معادلة نانلي Nunnally للدرجات الحقيقية.

*** تقديرات قدرات الأفراد Ability estimation:**

وتشير إلى مواقع الأفراد على متصل وتدريج الصفة المقاسة من الاختبار وتحدد امتلاكهم لمستويات محددة من هذه الصفة (Jason, Edward & Walter, 2009).

وتعرف إجرائيا بتقديرات قدرات الأفراد المشتقة من تحليلات كل من نموذجي التحليل (ثلاثي المعلم، التقدير الجزئي).

*** نوع المفردة Item type:**

ويعرف بنوع صياغة المفردة (اختيار من متعدد- صوابوخطأ- مزاوجة- تكملة- مقال....) وسوف تقتصر الدراسة على انواع محددة من الأسئلة وهي من نوع الاختيار من متعدد والصواب والخطأ والإكمال والمقال.

**إجراءات الدراسة:**

سوف تتبع الدراسة في ختبار الفرضيات الخاصة بها الإجراءات التالية:

**- اختيار العينة:**

تتكون عينة الدراسة من عدد (٧٢٢) بواقع عدد (٥٣٢) من طلاب وطالبات المستوى السابع بقسم علم النفس بكلية التربية بجامعة الملك سعود بالمستوى الدراسي الخامس (٣١٣ طالب + ٣١٩ طالبة)، وهي تمثل كل الدارسين الذي سبق لهم دراسة مقرري إحصاء (١) وإحصاء (٢). وعدد (١٩٠) من طلاب وطالبات الدراسات العليا من مستوى الماجيستير بواقع (١٢٤ طالب + ٦٦ طالبة) الدراسين لمقرر ٥٠٣ نفس (إحصاء تربوي)، ومقرري الإحصاء النفسي لطلاب الماجيستير تخصص علم النفس (٥٠٥ نفس، ٥٨٣ نفس).

**- أداة الدراسة:**

تكونت اداة الدراسة من صورتها المبدئية (١٨٠) مفردة اختبارية تم صياغتها بستة صور مختلفة وهي: (٣٠) مفردة من نوع الاختيار من متعدد + (٣٠) مفردة من نوع الصواب والخطأ + (٣٠) مفردة من نوع الإكمال + (٣٠) مفردة من نوع الإجابة القصيرة + (٣٠) مفردة من نوع المزاوجة + (٣٠) مفردات من نوع المقال، وقد صممت لقياس المهارات الإحصائية لدى طلاب المستوى السابع وطلاب الدراسات العليا (ماجيستير) والتي تم اكتسابها من خلال دراستهم لمقررات الإحصاء، وقد تم عرض الاختبار على مجموعة من المحكمين في مجال الإحصاء بهدف إبداء مرئياتهم في ضوء بعض المتغيرات التي تتعلق بعينة الأسئلة المحتواة بالأداة ومن بين هذه المتغيرات (سلامة المفردة من الناحية اللغوية والفنية، المستوى المعرفي التي تقيسه المفردة في ضوء تصنيف بلوم للأهداف، مدى تمثيل المفردات للموضوعات المستهدفة من الاختبار، وبعد الحصول على مرئيات المحكمين تم تعديل بعض الفقرات وإعادة

تصنيف بعض الفقرات في مستويات أخرى من تصنيف بلوم، وقد أسفرت التعديلات عن توافر رصيد من المفردات يتكون من (١٨٠) مفردة اختبارية، وقد تم وضع هذه المفردات في صورة اختبارية واحدة بهدف تحليلها وتدريجها كي تشكل بنكاً للأسئلة وذلك بتطبيق تلك الصورة الاختبارية على عينة من طلاب وطالبات كلية التربية بجامعة الملك سعود من مرحلتي البكالوريوس والماجيستير تكونت من (٧٢٢) فرد (على جلستين مختلفتين) وذلك إعتمادا على استنتخدام نموذج بارشال كريدت (الصورة العامة من نموذج راش) حيث يصلح استخدام هذا النموذج مع كل من المفردات الموضوعية ثنائية التقسيم والمفردات المقالية متعددة التقسيم، حيث يتعامل النموذج مع المفردات ثنائية التقسيم كحالة خاصة من المفردات متعددة التقسيم ( & Lihua Yao Schwarz, (2006: 475 وقد استخدم في التحليلات أيضاً النموذج الثلاثي البارمتر للحصول على بعض المؤشرات الإحصائية التي لا يمكن الحصول عليها من تحليلات نموذج التقدير الجزئي مثل متغيرات التمييز والتخمين.

وقد اسفرت التحليلات عن ملاءمة كل المفردات الاختبارية (١٨٠) مفردة في ضوء مؤشري الملاءمة (Infit، Outfit) المتضمنان ببرنامج Winsteps وعدم ملاءمة عدد (١١) فرد في ضوء نفس المؤشرين السابقين وبعد استبعاد الأفراد غير الملائمين وإعادة التحليل مرة اخرى تم الحصول على كل الخصائص والمؤشرات الإحصائية الخاصة بمفردات بنك الأسئلة وهي (مستوى الصعوبة، قوة التمييز ، مستوى التخمين، متوسط أقصى معلومات) وقد تألف الاختبار بصورته النهائية من عدد قدره (١٤٤) مفردة اختبارية تم سحبها من البنك المصغر للأسئلة والمكون من

(١٨٠ مفردة) ووضعها باداة الدراسة في ضوء خصائصها الإحصائية والمعرفية وفي ضوء ضوابط المعالجات البحثية الخاصة بالدراسة.

وقد روعي عند تصميم هذه الأداة مايلي:

- سحب المفردات البنكية التي تتدرج تحت نفس المواضيع الإحصائية بصيغ مختلفة (اختيار من متعدد + صواب وخطأ + الإكمال + الإجابة القصيرة + المزاوجة + المقال).

- توزيع مفردات الاختبار المصاغة بالأنواع الستة المذكورة آنفا على أبعاد اختبار الدراسة والتي تمثل نوع المفردات، وقد روعي تكافؤ المفردات داخل كل بعد في كل متغير من متغيرات الدراسة الأخرى والتي تتمثل في الصعوبة والتمييز والتخمين وأقصى معلومات وتمثيل نطاق المتغير المقاس والمستوى المعرفي.

- تكرار الخطوة السابقة مع متغير المستوى المعرفي تبعا لمستويات تصنيف بلوم وهي: التذكر، الفهم، التطبيق، عليا (تحليل، تركيب، تقويم).

وقد روعي تكافؤ المفردات داخل كل مستوى من مستويات الأهداف في كل من الصعوبة والتمييز والتخمين وأقصى معلومات وتمثيل النطاق ونوع السؤال.

- تقسيم عينة المفردات الكلية في ضوء تقديرات الصعوبة الخاصة بعد ترتيبها تصاعديا، من خلال درجتين للقطع في ثلاث فئات متساوية تختلف فيما بينها من حيث مستوى الصعوبة (منخفض، متوسط، مرتفع) حيث تحتوي كل فئة من الفئات الثلاث نسبة (٣٣.٣%) من

العدد الكلي للمفردات وبالتالي فإن الفئة الأولى تتضمن (٣٣.٣%) من أسهل المفردات، وتتضمن الفئة التالية لها نسبة (٣٣.٣%) من المفردات متوسطة الصعوبة وهي التي تلي مفردات الفئة الأولى في التدريج وتتضمن الفئة الثالثة النسبة المتبقية من المفردات وهي نسبة (٣٣.٣%) من المفردات مرتفعة الصعوبة وهي النسبة التي تلي الفئة الثانية مباشرة في تدريج الصعوبة الكلي، بحيث تتكافأ جميع المفردات المتناظرة داخل كل فئة من الفئات الثلاث في متغيرات التمييز والتخمين وأقصى معلومات ونوع المفردة وعدد المفردات والمستوى المعرفي ونسبة تمثيل النطاق.

- وقد تكرر الإجراء السابق مع متغير التمييز (بعد تحليل جميع المفردات باستخدام النموذج الثلاثي البارمتر) بما في ذلك أسئلة المقال بعد تحويل كل سؤال مقالي إلى عدد من الأسئلة الفرعية (ثنائية التقسيم) حيث يقابل كل سؤال فرعي خطوة من خطوات الإجابة المتعلقة بالسؤال نفسه–تقدر الإجابة الصحيحية بالدرجة (١) والدرجة (٠) إذا كانت خاطئة– وقد تم حساب مستوى التمييز للسؤال المقالي بمتوسط التمييز المقدر للأسئلة الفرعية التي تنبثق من السؤال.
- تكرار الإجراء السابق مع متغير التخمين، وذلك للأسئلة الموضوعية فقط دون المقالية وأسئلة الإجابة القصيرة حيث أعتبر أن مستوى التخمين المقدر لكل من هاتين النوعيتن من الأسئلة مساوٍ للصفر مع مراعاة تكافؤ الصور الثلاثة من حيث عدد الأسئلة ونوعها ومستوياتها المعرفية ومستويات الصعوبة والتمييز وتختلف من حيث مدى تمثيلها للنطاق السلوكي المستهدف من القياس.

- تقسيم الاختبار الكلي إلى ثلاث صور مختلفة ومتساوية من حيث عدد الأسئلة ومتكافئة من حيث نوعها ومستوياتها المعرفية والصعوبة والتمييز والتخمين وتختلف من حيث مدى تمثيلها لنطاق المتغير المستهدف من القياس:

- (الصورة الأولى) وتتمثل في الاختبار الكلي المتضمن لكل الأسئلة الواردة في كل الموضوعات المستهدفة من الاختبار.
- (الصورة الثانية) تحتوي على عدد من الأسئلة ترد في (٣/٢) ثلثي المواضيع المستهدفة من الاختبار الكلي (بنسبة مئوية (٦٦.٦%) تقريباً.
- (الصورة الثالثة) تحتوي على عدد من الأسئلة ترد في (٣/١) ثلث الموضوعات المستهدفة من الاختبار الكلي بنسبة مئوية (٣٣.٣%) تقريباً.

- تقسيم الاختبار إلى ثلاث صور تختلف من حيث أقصى معلومات (مستوى منخفض، مستوى متوسط، مستوى مرتفع) مع مراعاة تكافؤ الصور الثلاثة من حيث عدد الأسئلة ونوعها ومستوياتها المعرفية ومستويات الصعوبة والتمييز والتخمين تمثيلها للنطاق وتختلف من حيث أقصى معلومات.

والجدول التالي يوضح توزيع مفردات الاختبار الكلي على متغيرات الدراسة والذي يوضح إجراءات ضبط المتغيرات الأخرى في حال معالجة واحد من تلك المتغيرات.

# جدول (١)

## تصنيف مفردات اختبار الدراسة على مستويات المتغيرات البحثية

| الاختبار الكلي | | | | | | | | متغيرات تكافؤ أسئلة الاختبار | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| نوع السؤال | اختيار من متعدد | صواب وخطأ | مزاوجة | أكمال | إجابات قصيرة | مقال | المجموع | ن | ص | ت | خ | م | ف | ث |
| عدد الأسئلة | ٢٨ | ٢٨ | ٢٨ | ٢٨ | ١٢ | ١٢ | ١٤٤ | × | √ | √ | √ | √ | √ | √ |
| مستوى الصعوبة (ص) | سهل | متوسط | صعب | | | | | | | | | | | |
| عدد الأسئلة | ٤٨ | ٤٨ | ٤٨ | | | | ١٤٤ | √ | × | √ | √ | √ | √ | √ |
| مستوى التمييز (ت) | منخفض | متوسط | مرتفع | | | | | | | | | | | |
| عدد الأسئلة | ٤٨ | ٤٨ | ٤٨ | | | | ١٤٤ | √ | √ | × | √ | √ | √ | √ |
| مستوى التخمين (خ) | منخفض | متوسط | مرتفع | | | | | | | | | | | |
| عدد الأسئلة | ٤٨ | ٤٨ | ٤٨ | | | | ١٤٤ | √ | √ | √ | × | √ | √ | √ |
| أقصى معلومات (م) | منخفض | متوسط | مرتفع | | | | | | | | | | | |
| عدد الأسئلة | ٤٨ | ٤٨ | ٤٨ | | | | ١٤٤ | √ | √ | √ | √ | × | √ | √ |
| المستوى المعرفي (ف) | تذكر | فهم | تطبيق | عليا | | | | | | | | | | |
| عدد الأسئلة | ٣٨ | ٣٨ | ٣٨ | ٣٨ | | | ١٤٤ | √ | √ | √ | √ | √ | × | √ |
| تمثيل النطاق (ث) | ٣٣.٣% | ٦٦.٦% | ١٠٠% | | | | | | | | | | | |
| عددالأسئلة | ٤٨ | ٤٨ | ٤٨ | | | | ١٤٤ | √ | √ | √ | √ | √ | √ | × |

**حيث (ن) نوع السؤال، (ص) مستوى الصعوبة، (ت) مستوى التمييز، (خ) مستوى التخمين، (ن) أقصى معلومات، (ف) المستوى المعرفي، (ث) تمثيل النطاق.**

**التحقق من الخصائص السيكومترية للأداة:**

لقد تم التحقق من ثبات أداة الدراسة باستخدام معامل ألفا كرونباخ لتحقق من صدق الاتساق الداخلي للاختبار، وقد بلغ صدق الاتساق الداخلي للاختبار القيمة (٠.٩٤٨) وهي قيمة مرتفعة ونظرا لأن أداة الدراسة هي اختباراً تحصيلياً وبالتالي فإنه لا يقيس تكوين أو تكوينات فرضية قد يختلف تعريفها من فرد أو باحث لآخر فقد اكتفى الباحث بالتعرف على صدق الاختبار من خلال عرضه على مجموعة من المحكمين المتخصصين في مجال الاختبار (الإحصاء) وقد استخدم الباحثة النسبة (٨٥%) من موافقة المحكمين على سلامة مفردات الاختبار وهي تلك النسبة التي تحققت مع جميع فقرات الاختبار بصورته النهائية، كما أجرى الباحث استخدام الصدق الداخلي للأداة كما أشارت "انستازي" Anastasi إلى أن صدق الفقرات يمكن أن يحسب من خلال ارتباطها بمحك خارجي او داخلي وفي حالة عدم توافر محك خارجي فإن أفضل محك داخلي هو الدرجة الكلية للمقياس (Anastasi, 1988: 211) وقد قام الباحث بالفعل كما أشارت "أنستازي" إلى إيجاد معامل إرتباط الدرجة الخاصة بكل فقرة مع الدرجة الكلية وجاءت كل معاملات الارتباط طردية دالة عند مستوى (٠.٠١).

**خطة التحليل الإحصائي:**

سوف تعتمد الدراسة في إجراءات اختبار فرضياتها على بعض الأساليب الإحصائية وبيانها فيما يلي:

- النموذج اللوغاريتمي ثلاثي المعلم، وسوف يستخدم هذا النموذج في تحليل مفردات أداة الدراسة وذلك باستخدام بنامج بايلوج، BILOG-

MG حيث يتم من خلال تحليلات هذ النموذج التوصل على عدد من المؤشرات الإحصائية للمفردات المستهدفة من الأداة وهي (Zimowski, Muraki, Mislevy & Bock, 2007): (مؤشر الصعوبة– مؤشر التمييز – مؤشر التخمين– مؤشر المعلومات) لكل نوع المفردة من انواع المفردات (اختيار من متعدد– إكمال– مزاوجة...) وذلك لمفردات الأداة من النوع الموضوعي (تنائي التقسيم).

- نموذج التقدير الجزئي اللوغاريتمي (نموذج ماسترز MASTERS MODEL)، وذلك لتقدير مؤشر الصعوبة والمعلومات وذلك باستخدام برنامج WINSTEPS، وذلك لمفردات الأداة من النوع المقالي (متعددة التقسيم) (Liancer, 2005: 197).
- التحليل التمييزي Discriminant Analysis.
- معامل كابا Kappa Coefficient.
- معامل إرتباط الرتب لسبيرمان Spearman Coefficient.
- معادلة نانلي لتقدير الدرجة الحقيقية True score، وهي: T = R (X-M)+M:(Nunnally, & Bernstein,1994: 56).
- حيث (T) الدرجة الحقيقية، (M) متوسط الدرجات، (R) معامل ثبات الاختبار، (X) الدرجة الملاحظة.
- برامج التحليل الإحصائي المستخدمة:

– برنامج–BILOG MG3، (ب) برنامج WINSTEPS.

– الإجر اءات الخاصة باختبار فرضيات الدراسة.

- تطبيق أداة الدراسة التي تحتوي على عدد (١٤٤) مفردة اختبارية على عينة الدراسة.

- تقدير مستويات القدرة لجميع أفراد عينة الدراسة وذلك بتحليل استجابات عينة الأفراد على جميع مفردات الأداة (بعد دمجها بملف واحد) باستخدام تحليلات نموذج التقدير الجزئي ببرنامج WINSTEPS والذي يتيح تحليل كل الاستجابات على كل من نوعي المفردات (موضوعية ومقالية).
- تقسيم عينة الدراسة بعد ترتيبهم، وذلك وفقا لمعدلهم التراكمي GPA (Grade Point Average) إلى مجموعتين من المنخفضين والمرتفعين باستخدام أعلى نسبة (٣٣%) من مرتفعي المعدل التراكمي وادنى نسبة (٣٣%) من المعدل التراكمي.
- إيجاد مجموع درجات عينة الدراسة الكلية على الأسئلة التي تمثل كل مستوى من مستويات متغير الصعوبة (الأسئلة السهلة) وتمثل المستوى السهل، وكذلك الأسئلة متوسطة الصعوبة وتمثل (المستوى المتوسط من الصعوبة) وكذلك (الأسئلة الصعبة) وتمثل المستوى المرتفع من الصعوبة.
- تكرار الإجراء السابق بالخطوة (٤) مع الأسئلة التي تندرج تحت كل مستوى من المستويات الثلاث (المنخفض، المتوسط، المرتفع) للمتغيرات البحثية الأخرى (التمييز، أقصى معلومات، التخمين، تمثيل نطاق المتغير المقاس).
- إيجاد مجموع درجات عينة الدراسة لكل نوع من انواع أسئلة البحث (اختيار من متعدد، صواب وخطأ، مزاوجة، إكمال، إجابة قصيرة، مقال).
- إيجاد مجموع درجات عينة الدراسة للأسئلة التي تندرج تحت كل مستوى من المستويات المعرفية بتصنيف بلوم للأهداف (تذكر، فهم، تطبيق، عليا).

- استخدام التحليلات الإحصائية الخاصة بأسلوب التحليل التمييزي للمتغير التابع التصنيفي وهو متغير مرتفعي ومنخفضي المعدل التراكمي والمستويات المحددة لكل متغير من متغيرات الدراسة باعتبارها متغيرات مستقلة (تمييزية)..
- إيجاد معامل إرتباط الرتب لسبيرمان بين تقديرات القدرة لأفراد عينة الدراسة والتي اشتقت من تحليلات بيانات الاختبار الكلي باستخدام نموذج التقدير الجزئي (بعد تحويلها من وحدة اللوجيت إلى وحدة الوات) والمعدل التراكمي لنفس الأفراد.
- كلما إزدادات قيمة معامل إرتباط الرتب بين تقديرات القدرة لأفراد عينة الدراسة، والمعدل التراكمي لنفس الأفراد عند مستوى محدد من مستويات أي متغير بحثي كلما دل ذلك كفاءة هذا المستوى واعتباره مؤشرا يمكن الاعتماد عليه في تقييم الكفاءة النسبية الاختبار
- كلما إزدادت القيمة المطلقة للمعامل المعياري لأي متغير تمييزي (شامل، ٢٠٠٩، ٦٥٥)،(يمثل مستوى محدد من مستويات المتغير البحثي) كلما دل ذلك على درجة اسهام هذا المتغير التمييزي في التمييز بين عينتي المرتفعين والمنخفضين في المعدل التراكمي، ودل ذلك على كفاءة هذا المستوى واعتباره مؤشرا يمكن الاعتماد عليه في تقييم الكفاءة النسبية الاختبار.
- استخدام معادلة نانلي للدرجة الحقيقية لإيجاد الدرجة الحقيقية لكل فرد من أفراد عينة الدراسة على الأسئلة التي تمثل كل مستوى من مستويات كل متغير من متغيرات الدراسة.

- إيجاد معامل إرتباط الرتب لسبيرمان بين الدرجات الحقيقية لعينة الدراسة لكل مستوى من مستويات المتغيرات البحثية والمعدل التراكمي لنفس الأفراد.

- كلما إزداد معامل إرتباط الرتب بين بين الدرجات الحقيقية والمعدل التراكمي لنفس الأفراد عند مستوى محدد من مستويات أي متغير بحثي كلما دل ذلك كفاءة هذا المستوى واعتباره مؤشرا يمكن الاعتماد عليه في تقييم الكفاءة النسبية الاختبار.

- إشتقاق معايير لتقييم الكفاءة النسبية للاختبار التحصيلي البنكي من خلال تحديد درجات موزونة لقيم المعاملات المعيارية للمتغيرات التمييزية، وايضا معاملات إرتباط الرتب.

## عرض ومناقشة نتائج الدراسة:

**أسفرت إجراءات الإجابة عن أسئلة الدراسة عن النتائج التالية:**

**أولاً: نتائج الإجابة عن السؤال الأول:**

أسفرت إجراءات الإجابة عن السؤال الأول والذي ينص على: ما دلالة القوة التمييزية لكل متغير من متغيرات (الصعوبة، التمييز، أقصى معلومات، التخمين) بين مرتفعي ومنخفضي المعدل التراكمي التحصيلي؟ عن النتائج التالية:

### نتائج التحليل التمييزي لمتغير لمستوى صعوبة المفردات:

أسفرت التحليلات التي اعتمدت على المستويات الثلاث من متغير مستوى الصعوبة (منخفض- متوسط- مرتفع) (كتغيرات ثلاثة مستقلة

تمييزية) ومتغير مجموعتي مرتفعي ومنخفضي المعدل التراكمي (كتغير تابع تصنيفي) عن النتائج الموضحة بالجداول التالية:

**جدول (٢)**

**جدول الإحصاءات الوصفية لمتغيرات أقصى معلومات**

**كما يبدو من نتائج التحليل التمييزي**

| المجموعة | مستويات متغير الصعوبة | المتوسط | الإنحراف المعياري |
|---|---|---|---|
| منخفضين | منخفض | ٢١.٦٤٨٩ | ٤.٩٧١١٨ |
| | متوسط | ١٦.٩٨٩٤ | ٥.٣٥٢٦٨ |
| | مرتفع | ٨.١٧٠٢ | ٣.٩٩٩٠٣ |
| مرتفعين | منخفض | ٢٧.٦٠٢٧ | ١.٤٦٤٤٣ |
| | متوسط | ٢٦.٣٨٣٦ | ٢.٢٩٦٢٠ |
| | مرتفع | ٢٠.٣٢٨٨ | ٥.٢٠٦٢٤ |

يتبين من النتائج الموضحة بالجدول السابق (٢) وجود فروق ظاهرية بين متوسطات درجات عينتي منخفضي ومرتفعي المعدل التراكمي على أنواع الأسئلة التي تختلف باختلاف مستوى الصعوبة (منخفض، متوسط، مرتفع)، بلغ أعلى متوسط القيمة (٢٦.٣٨٣٦)، وكان لدرجات عينة مرتفعي المعدل التراكمي على الأسئلة متوسطة الصعوبة، بينما بلغ أدنى متوسط القيمة (٨.١٧٠٢) وكان لعينة منخفضي المعدل التراكمي على الأسئلة مرتفعة الصعوبة.

**جدول (٣)**

**جدول اختبار Box's M لدلالة تجانس تباين أو تغاير مجموعتي المرتفعين والمنخفضين**

| | |
|---|---|
| اختبار Box's M | ٦.٨٢٢ |
| مستوى الدلالة | ٠.١٠١ |

ويؤكد الجدول السابق (٣)، أن قيمة (ف) لاختبار (Box's M) بلغت (٦.٨٢٢) وهي قيمة غير دالة، وبالتالي فإن البيانات تستوفي شرط الاعتدالية وتجانس التباين والتغاير.

**جدول (٤)**

**معلومات حول خطوات دخول المتغيرات ودلالة كل منها في التحليل التمييزي**

| | الخطوة | اختبار Tolerance | Sig. of F to Remove | Wilks' Lambda |
|---|---|---|---|---|
| ٣ | المستوى الصعب | ٠.٧٦٧ | ٠.٠٠٠ | ٠.٤٩٤ |
| | المستوى متوسط | ٠.٢٩٥ | ٠.٠٠٠ | ٠.٣٧١ |
| | المستوى السهل | ٠.٢٦١ | ٠.٠٠٠ | ٠.٣٣٤ |

يتضح من الجدول (٤) من الخطوة الأخيرة رقم (٣) عدم تجاوز أي متغير النسبة (٠.١٠) المحددة لدلالة الخروج من نموذج التحليل Sig. of F to Remove، ورتبت المتغيرات وفقاً لقيمة اختبار "ويلكس لامدا" لنسبة التباين بين المجموعتين الذي يفسره كل منها؛ حيث يتم البدء بالمتغير الأكثر تأثيراً ثم المستوى المتوسط ثم المستوى المنخفض.

## جدول (٥)

## قيمة الجذر الكامن ومعامل الارتباط القانوني

| الدالة | الجذر الكامن | نسبة التباين % | النسبة التراكمية للتباين % | معامل الإرتباط القانوني | مربع إيتا |
|---|---|---|---|---|---|
| ١ | a٢.٤١٩ | ١٠٠ | ١٠٠ | ٠.٨٤١ | ٠.٧٠٧ |

يتضح من الجدول (٥) أن قيمة الجذر الكامن بلغت (٢.٤١٩) كما أن قيمة الارتباط القانوني والذي يسمى أحياناً الارتباط الطبيعي أو التجميعي قد بلغت (٠.٨٤١)، وهو يشير إلى مدى الارتباط بين الدرجات التمييزية وفئات المتغير النوعي أو المجموعات. وبتربيع قيمة هذا المعامل تم الحصول على قيمة مربع إيتا التي بلغت (٠.٧٠٧)، تقريبا وهو أحد مؤشرات حجم الأثر التي تعتمد على تحديد نسبة التباين المفسر.

## جدول (٦)

## قيمة معامل ويلكس لامدا ودلالته الإحصائية

| اختبار الدوال | اختبار قيمة Wilks' Lambda | مربع كاي | درجات الحرية | مستوى الدلالة |
|---|---|---|---|---|
| ١ | ٠.٢٩٢ | ٥٢١.٨٦٤ | ٣ | ٠.٠٠٠ |

ويبين الجدول (٦) أن قيمة معامل ويلكس لامبدا بلغت (٠.٢٩٢)، وهي دالة عند مستوى (٠.٠١). ومعامل ويلكس لامبدا هو مؤشر على كمية التباين غير المفسر في الدرجات التمييزية، وكلما صغرت قيمة هذا

المعامل كانت نتائج التحليل أفضل؛ حيث تدل القيمة الأقرب إلى الصفر على اختلاف المجموعات.

**جدول (٧)**

**المعاملات المعيارية للدالة التمييزية**

| مستويات متغير صعوبة المفردات | الدالة |
|---|---|
| | ١ |
| المستوى السهل | -٠.٨١٦ |
| المستوى متوسط الصعوبة | ١.٠٠٨ |
| المستوى الصعب | ٠.٨٠٧ |

ويتضح من الجدول (٧) الذي يوضح معاملات المعيارية للدالة التمييزية، أن القيم الواردة للمعاملات تبين أن المستوى المتوسط من متغير مستوى صعوبة المفردات بلغ معامله المعياري القيمة (١.٠٠٨) وهو اكبر من المعامل المعياري لمتغير المستوى المرتفع من مستوى صعوبة المفردات والذي بلغت قيمته (٠.٨٦٧) وبالتالي يتضح أن المستوى المتوسط من مستوى الصعوبة كان مؤثرا في النموذج بدرجة تفوق تأثير المستوى المرتفع من نفس المتغير.

## جدول (٨)

## معامـلات الدالة التمييزيــة

| مستويات متغير صعوبة المفردات | الدالة |
|---|---|
| | ١ |
| المستوى السهل | ٠.١٩٨- |
| المستوى متوسط الصعوبة | ٠.٢٢٢ |
| المستوى الصعب | ٠.١٩٥ |
| الثابت | ٢.١٩١- |

يتبين من الجدول (٨) معاملات الدالة التمييزية لكل متغير من المتغيرات التمييزية (المستقلة) وهذه القيم هي قيم (أ) للمستوى السهل من متغير مستوى صعوبة المفردات بلغت قيمتها (-٠.١٩٨)، و(ب) للمستوى المتوسط من متغير مستوى الصعوبة وبلغت قيمته (٠.٢٢٢)، و(ج) للمستوى الصعب من متغير مستوى الصعوبة وبلغت قيمته (٠.١٩٥) بالإضافة على ثابت المعادلة (ث) والذي بلغت قيمته (-٢.١٩١) وهي القيم اللازمة لمعادلة الانحدار التي تمكننا من التنبؤ بالمجموعة التي تنتمي إليها الحالات الجديدة، كما سيرد لا حقاً.

## جدول (٩)

## إقتران تصنيف عضوية أفراد فئتي منخفضي ومرتفعي التراكمي بالتصنيف المتنبا به من نموذج التحليل

| المجموعة الأصلية المصنفة في ضوء المعدل التراكمي | المجموعة التنبأ بها من التحليل | | المجموع |
|---|---|---|---|
| | منخفضين | مرتفعين | |
| منخفضين | ٢٧٩ | ٣ | ٢٨٢ |
| مرتفعين | ١٨ | ٢٢٨ | ٢٤٦ |

من خلال جدول (٩) يتضح أن إحدى وعشرون حالة فقط (٣ + ١٨ = ٢١) كان التنبؤ بها خاطئاً حيث تم تصنيفهم من قبل التحليل في مجموعات خاطئة فقد صنفت (١٨) حالة من حالات من مرتفعي المعدل التراكمي بفئة منخفضي المعدل التراكمي، بينما صنفت (٣) حالات من منخفضي المعدل التراكمي بفئة مرتفعي المعدل الراكمي، بينما بقية الحالات (٢٧٩ + ٢٢٨ = ٥٠٧) كان التنبؤ بها صحيحاً، وهذا يشير إلى أن نسبة صحة التنبؤ تبلغ (٥٠٧ ÷ ٥٢٨) = ٩٦% تقريباً.

**جدول (١٠)**

**معامل قوة الاتفاق لكابا بين التصنيف المتنبأ به والتصنيف الأصلي لأفراد فئتي منخفضي ومرتفعي المعدل التراكمي**

| معامل كابا Kappa | القيمة | مستوى الدلالة |
|---|---|---|
| | ٠.٩٢٦ | ٠.٠٠٠ |

وتؤكد قيمة معامل كابا على فاعلية النموذج في التنبؤ، فقد بلغت (٠.٩٣) تقريبا مما يدل على وجود ارتباط قوي بين النتيجة الفعلية والنتيجة المتنبأ بها، مما يشير إلى قوة اتساق التصنيف المتوقع من النموذج الإحصائي لأفراد عينة البحث مع تصنيفهم الأصلي بمجموعتي منخفضي ومرتفعي المعدل التراكمي إعتماد على المتغيرات التمييزية المتضمنة به والممثلة بالمستويات الثلاثة من متغير الصعوبة (منخفض، متوسط، مرتفع).

يتضح مما سبق من نتائج أن الأسئلة الاختبارية التي تمتلك قدرا متوسطا من الصعوبة تفوقت على الأسئلة الأخرى التي امتلكت قدرا مرتفعا ومنخفضا من الصعوبة حيث كان للأسئلة متوسطة الصعوبة الأثر الأكبر في نموذج التحليل، وبالتالي كانت أكثر قدرة على التمييز

بين عينتي مرتفعي ومنخفضي المعدل التراكمي، يلي ذلك الأسئلة الصعبة ثم السهلة بما يفيد بأن الأسئلة التي تمتلك المستوى المتوسط من الصعوبة يمكن الاعتماد عليه في تقييم الكفاءة النسبية للاختبار ويمكن أن يكون لها وزنا أعلى في تقييم الاختبار عن أوزان الأسئلة الأخرى الصعبة ثم السهلة، وتؤكد تلك النتائج ماورد بالتراث فيما يتعلق بكفاءة الأسئلة متوسطة المستوى من حيث الصعوبة في دقة تقديرات الاختبار لمستويات القدرة المستهدفة من القياس وأيضا دقة التمييز بين المستويات المختلفة من القدرة (فرج، ٢٠٠٩، ٣٢٢)، (مصطفى، ٢٠٠٠)، كما تبين من أيضاً من خلال التراث أن الأسئلة الصعبة يمكن أن يكون لها نصيباً من أسئلة الاختبار باعتبار أنها تميز الطلاب المتفوقون عن زملائهم العاديون.

فقد أظهرت دراسة (Aaron, Irina & David, 2005) أن الأسئلة التي تقيس الممستويات العقلية والمعرفية العليا هي أسئلة صعبة وتشير إلى كفاءة الاختبار، وعلى الرغم من صعوبة تلك الأسئلة إلا أنها ترتبط بالتقويم الحقيقي لعملية التعلم وبالتالي ينبغي أن يكون لها حظاً من أسئلة تلك الاختبارات، وهذا ما تأكد أيضا من خلال دراسة Kucuk & Walters(2009) التي اشارت إلى بعض المؤشرات الإحصائية التي يمكن الاستناد إليها في قياس كفاءة الاختبار مثل مؤشرات الصعوبة التي تفيد بوجوب تمتع أسئلة الاختبار بمستوى متوسط من الصعوبة وبعض الأسئلة مرتفعة الصعوبة، وتؤكد هذه النتيجة ما أشارت إليه بعض الاطر النظرية التي تفيد بضرورة اتساع مدى صعوبة أسئلة الاختبار حتى يتصف الاختبار بحساسية تمييزه للمستويات المختلفة والمتعددة من القدرة المقاسة (Anastasi, 1988: 174)، كما أن وجود نسبة من

الأسئلة الصعبة بدورة لايؤثر على الخصائص السيكومترية للاختبار والتي تعتبر من أهم المؤشرات الدالة على جودة الاختبار وكفاءته، وقد أشارت بعض الدراسات مثل دراسة Jandaghi & Shateria (2008) إلى عدم تأثر معاملات صدق وثبات الاختبار بوجود نسبة قليلة من الأسئلة مرتفعة الصعوبة، وتبدو تلك النتيجة منطقية نظرا لأن مستوى صعوبة أسئلة الاختبار يعتبر من بين المتغيرات التي ترتبط بالدرجة الحقيقية التي يمكن أن يحصل عليها متلقي الاختبار فقد أكدت دراسة (مصطفى، ٢٠٠٠) التي أظهرت وجود فروق ذات دلالة إحصائية للدرجة الحقيقية للطالب بين مستويات الصعوبة (سهل- متوسط- صعب) في صالح المستويين السهل والمتوسط لأسئلة الاختيار من متعدد وكذلك مع أسئلة الصواب والخطأ دون المستوى الصعب من الأسئلة وهذا ما يتعارض مع الدراسة الحالية، وقد يعزى اختلاف هذه النتيجة مع النتيجة التي توصلت إليها الدراسة الحالية فيما يتعلق بالأسئلة الصعبة إلى تحديد درجة قطع لتحديد الأسئلة مرتفعة الصعوبة بدراسة (مصطفى، ٢٠٠٠) عن درجة القطع التي استخدمت في تحديد نفس الأسئلة بالدراسة الحالية، كما أن النتائج التي توصلت إليها الدراسة الحالية والتي تتعلق بتفوق قدرة المفردات التي تمتلك مستوى متوسط من الصعوبة في التمييز بين الأفراد مرتفعي ومنخفضي المعدل التراكمي اكثر من قدرة المفردات الصعبة والسهلة تؤكد ما أوضحته دراسة Tavakol, et.al., (2011) من أن بعض المؤشرات يمكن أن يعتمد عليها في تحديد كفاءة الاختبار التي تبدو في المستوى المتوسط من صعوبة بنود الاختبار.

**نتائج التحليل التمييزي لمتغير تمييز المفردات:**

أسفرت التحليلات التي اعتمدت على المستويات الثلاث من متغير التمييز (منخفض- متوسط- مرتفع) (كتغيرات ثلاثة مستقلة تمييزية) ومتغير مجموعتي مرتفعي ومنخفضي المعدل التراكمي (كتغير تابع تصنيفي) عن النتائج الموضحة بالجداول التالية:

**جدول (١١)**

**جدول الإحصاءات الوصفية لمتغيرات أقصى معلومات كما يبدو من نتائج التحليل التمييزي**

| المجموعة | مستويات متغير التمييز | المتوسط | الإنحراف المعياري |
|---|---|---|---|
| منخفضين | المستوى المنخفض | ١٥.٣٥٤٦ | ٤.٨١٨٦١ |
| | المستوى المتوسط | ٥.٧٠٩٢ | ١.٣٥٠٠٤ |
| | المستوى المرتفع | ١٤.٩٩٢٩ | ٣.٧٤٢٦ |
| مرتفعين | المستوى المنخفض | ٢٤.٠٩٥٩ | ٢.٢١٣٨٤ |
| | المستوى المتوسط | ٧.١٢٣٣ | ٠.٨٥٤١٢ |
| | المستوى المرتفع | ٢٣.٣٦٩٩ | ٢.٩٧٥٨ |

يتبين من النتائج الموضحة بجدول (١١) وجود فروق ظاهرية بين متوسطات درجات عينتي منخفضي ومرتفعي المعدل التراكمي على أنواع الأسئلة التي تختلف باختلاف قوة تمييز المفردات (منخفضة، متوسطة، مرتفعة)، بلغ أعلى متوسط القيمة (٢٤.٠٩٥٩)، وكان لدرجات عينة مرتفعي المعدل التراكمي على الأسئلة منخفضة التمييز، بينما بلغ أدنى متوسط القيمة (٥.٧٠٩٢) وكان لعينة منخفضي المعدل التراكمي على الأسئلة متوسطة التمييز.

**جدول (١٢)**

**جدول اختبار Box's M لدلالة تجانس تباين أو تغاير مجموعتي المرتفعين والمنخفضين**

| | |
|---|---|
| اختبار Box's M | ٣.٣٥٤ |
| مستوى الدلالة | ٠.٢٣٠ |

ويؤكد الجدول (١٢)، أن قيمة (ف) لاختبار (Box's M) بلغت (٣.٣٥٤) وهي قيمة غير دالة، وبالتالي فإن البيانات تستوفي شرط الاعتدالية وتجانس التباين والتغاير.

**جدول (١٣)**

**معلومات حول خطوات دخول المتغيرات ودلالة كل منها في التحليل التمييزي**

| الخطوة | | Tolerance | Sig.of F to Remove | Wilks' Lambda |
|---|---|---|---|---|
| ٣ | مستوى التمييز المرتفع | ٠.٤٦٣ | ٠.٠٠٠ | ٠.٥٤٩ |
| | مستوى التمييز المتوسط | ٠.٣٢٧ | ٠.٠٠٠ | ٠.٤١٩ |
| | مستوى التمييز المنخفض | ٠.٣٢٩ | ٠.٠٠٠ | ٠.٤١٧ |

يتضح من الجدول (١٣) من آخرخطوة (٣) عدم تجاوز أي متغير النسبة (٠.١٠) المحددة لدلالة الخروج من نموذج التحليل Sig. of F to Remove، ورتبت المتغيرات وفقاً لقيمة اختبار "ويلكس لامدا" لنسبة التباين بين المجموعتين الذي يفسره كل منها؛ حيث يتم البدء بالمتغير الأكثر تأثيراً ثم المستوى المتوسط ثم المستوى المنخفض.

**جدول (١٤)**

**قيمة الجذر الكامن ومعامل الارتباط القانوني**

| الدالة | الجذر الكامن | نسبة التباين % | النسبة التراكمية للتباين % | معامل الإرتباط القانوني | مربع إيتا |
|---|---|---|---|---|---|
| ١ | ١.٧٩٢ | ١٠٠ | ١٠٠ | ٠.٨٠١ | ٠.٦٤٢ |

يتضح من جدول (١٤) أن قيمة الجذر الكامن بلغت (١.٧٩٢) كما أن قيمة الارتباط القانوني قد بلغت (٠.٨٠١)، وهو يشير إلى مدى الارتباط بين الدرجات التمييزية وفئات المتغير النوعي أو المجموعات. وبتربيع قيمة هذا المعامل تم الحصول على قيمة مربع إيتا التي بلغت (٠.٦٤٢)، تقريبا وهو أحد مؤشرات حجم الأثر التي تعتمد على تحديد نسبة التباين المفسر.

**جدول (١٥)**

**قيمة معامل ويلكس لامدا ودلالته الإحصائية**

| اختبار الدوال | اختبار قيمة Wilks' Lambda | مربع كاي | درجات الحرية | مستوى الدلالة |
|---|---|---|---|---|
| ١ | ٠.٣٥٨ | ٤٣٥.٨٩٨ | ٣ | ٠.٠٠٠ |

ويبين جدول (١٥) أن قيمة معامل ويلكس لامبدا بلغت (٠.٣٥٨)، وهي دالة عند مستوى (٠.٠١). ومعامل ويلكس لامبدا هو مؤشر على كمية التباين غير المفسر في الدرجات التمييزية، وكلما صغرت قيمة هذا المعامل كانت نتائج التحليل أفضل؛ حيث تدل القيمة الأقرب إلى الصفر على اختلاف المجموعات.

## جدول (١٦)

## المعامـلات المعيارية للدالة التمييزيـة

| مستويات متغير التمييز | الدالة |
|---|---|
| | ١ |
| مستوى التمييز المنخفض | ٠.٨١٣ |
| مستوى التمييز المتوسط | ٠.٨٦١ |
| مستوى التمييز المرتفع | ٠.٨٣٣ |

ويتضح من جدول (١٦) الذي يحتوي المعاملات المعيارية للدالة التمييزية، أن المستوى المتوسط من متغير قوة تمييز المفردات بلغ معامله المعياري القيمة (٠.٨٦١) وهو اكبر من المعامل المعياري لمتغير المستوى المرتفع من نفس المتغير والذي بلغت قيمته (٠.٨٣٣)، وكذلك المعامل المعياري للمستوى المنخفض بلغت قيمته (٠.٨١٣) وبالتالي فإن المستوى المتوسط من متغير قوة تمييز المفردات كان مؤثراً في النموذج بدرجة تفوق تأثير المستوى المرتفع وكذلك المنخفض من نفس المتغير.

## جدول (١٧)

## معامــلات الدالة التمييزيــة

| مستويات متغير التمييز | المعاملات |
|---|---|
| مستوى التمييز المنخفض | ٠.١٩٩ |
| مستوى التمييز المتوسط | ٠.٦٩١ |
| مستوى التمييز المرتفع | ٠.٢٤٦ |
| الثابت | -٣.٧٤٩ |

يتبين من الجدول (١٧) معاملات الدالة التمييزية لكل متغير من المتغيرات التمييزية (المستقلة) وهذه القيم هي قيم (أ) للمستوى المنخفض من متغير قوة تمييز المفردات بلغت قيمتها (٠.١٩٩)، و(ب) للمستوى المتوسط من متغير قوة التمييز وبلغت قيمته (-٠.٦٩١)، و(ج) للمستوى الرمتفع من متغير قوة التمييز وبلغت قيمته (٠.٢٤٦) بالإضافة على ثابت المعادلة (ث) والذي بلغت قيمته (-٣.٧٤٩) وهي القيم اللازمة لمعادلة الانحدار التي تمكننا من التنبؤ بالمجموعة التي تنتمي إليها الحالات الجديدة.

**جدول (١٨)**

**إقتران تصنيف عضوية أفراد فئتي منخفضي ومرتفعي التراكمي بالتصنيف المتنبا به من نموذج التحليل**

| المجموعة الأصلية المصنفة في ضوء المعدل التراكمي | المجموعة التنبأ بها من التحليل | | المجموع |
|---|---|---|---|
| | منخفضين | منخفضين | |
| منخفضين | ٢٦٤ | ١٨ | ٢٨٢ |
| مرتفعين | ٤ | ٢٤٢ | ٢٤٦ |

يتضح من جدول (١٨) أن إثنى وعشرون حالة فقط (٤ + ١٨ = ٢٢) كان التنبؤ بها خاطئاً حيث تم تصنيفهم من قبل التحليل في مجموعات خاطئة.

فقد صنفت (١٨) حالات من منخفضي المعدل التراكمي بفئة مرتفعي المعدل التراكمي، بينما صنفت (٤) حالات من مرتفعي المعدل التراكمي بفئة منخفضي المعدل الراكمي، بينما بقية الحالات (٢٦٤ + ٢٤٢ = ٥٠٦) كان التنبؤ بها صحيحاً، وهذا يشير إلى أن نسبة صحة التنبؤ تبلغ (٥٠٦ ÷ ٥٢٨) = ٩٥.٨% تقريباً.

**جدول (١٩)**

**معامل قوة الاتفاق لكابا بين التصنيف المتنبأ به والتصنيف الأصلي لأفراد فئتي منخفضي ومرتفعي المعدل التراكمي**

| اختبار Kappa | القيمة | مستوى الدلالة |
|---|---|---|
| | ٠.٨٩٩ | ٠.٠٠٠ |

وتؤكد قيمة معامل كابا من جدول (١٩) على فاعلية النموذج في التنبؤ، فقد بلغت (٠.٩٠) تقريبا مما يدل على وجود ارتباط قوي بين النتيجة الفعلية والنتيجة المتنبأ بها. مما يشير إلى قوة اتساق التصنيف المتوقع من النموذج الإحصائي إعتماد على المتغيرات التمييزية المتضمنة به والممثلة بالمستويات الثلاثة من متغير التمييز (منخفض، متوسط، مرتفع) لأفراد عينة البحث مع تصنيفهم الأصلي بمجموعتي منخفضي ومرتفعي المعدل التراكمي.

يتضح مما سبق من نتائج أن الأسئلة الاختبارية التي تمتلك قدرا متوسطا من التمييز تفوقت على الأسئلة الأخرى التي امتلكت قدرا مرتفعا ومنخفضا من نفس المتغير حيث كان للأسئلة متوسطة التمييز الأثر الأكبر في نموذج التحليل وبالتالي كانت أكثر قدرة على التمييز بين عينتي مرتفعي ومنخفضي المعدل التراكمي، يلي ذلك الأسئلة مرتفعة التمييز ثم المنخفضة بما يفيد بأن الأسئلة التي تمتلك المستوى المتوسط من التمييز يمكن الاعتماد عليها في تقييم الكفاءة النسبية للاختبار ويمكن أن يكون لها وزنا أعلى في تقييم الاختبار عن أوزان الأسئلة الأخرى المرتفعة والمنخفضة من حيث التمييز، وقد أكدت هذه النتيجة بعض الدراسات التي استهدفت التعرف على بعض أنواع الأسئلة التي تتمتع بقدرة عالية على التمييز فقد أظهرت دراسة Clifton (2010)

Schriner &أن هذه الأسئلة التي تتمتع بمستوى متوسط وفقا لمؤشرات التمييز ، يكون لها القدرة على التمييز بين المستويات التحصيلية المختلفة.

كما أن المستوى المتوسط من التمييز لا يقتصر على نوعية معينة من المفردات دون الأخرى بما يتيح لواضع الاختبار اختيار انواع مختلفة من المفردات يمكن أن تتمتع بمستوى من التمييز يرغب به، كما أشارت دراسة Tasdemir & Mehmet (2010) إلى أن أسئلة الاختيار الصواب والخطأ والمزاوجة التي تتمتع بالمستوى المتوسط من متغير التمييز لها نفس الخصائص الخاصة بالتمييز التي تتمتع بها أسئلة الاختيار من متعدد.

ونظرا لأن قوة تمييز المفردات الاختبارية تلعب دورا هاما في التفريق بين المتعلمين في العديد من المهارات المستهدفة تعليميا، فقد أشارت بعض الأدبيات إلى ضرورة التمييز بين الطلاب في بعض المهارات اعتمادا على ادوات تتمتع بقدر مناسب من التمييز لا يميل إلى الإرتفاع أو الإنخفاض الشديد، حيث أوضحت دراسة Ryan, Peter & Elizabeth (2010) أن أكثر الأدلة صدقا في التمييز بين الطلاب تلك المرتبطة بالمهارات الأكاديمية ومهارات التفكير الناقد وكذلك المهارات الحسابية (الرياضية).

وتوحي نتائج تلك الدراسة أن المستوى المناسب وغير المتطرف من متغير التمييز يمكن الاستناد إليه كاحد المعايير التي يمكن استخدامها في تقييم الكفاءة النسبية للاختبار الذي ينبغي أن يميز بين الطلاب في المهارات المشار إليها، هذا وتؤكد النتائج على فاعلية متغير تمييز المفردة الاختبارية وبصفة خاصة المستوى المتوسط من هذا المتغير، في

تقدير مستوى الأداء الحقيقي للطالب نظرا لإرتباط قوة تمييز مفردات الاختبار بالدرجة الحقيقية الذي يمكن الحصول عليها من الاختبار.

وفي هذا الصدد أوضحت دراسة Tavakol, et.al., (2011) أن بعض المتغيرات تؤثر على الدرجة الحقيقية التي يمكن أن يحصل عليها الطالب في الاختبار من بينها مستوى تمييز بنود الاختبار وبصفة خاصة المستوى المتوسط منها، وهذا ما تبين من خلال نتائج الدراسة الحالية والتي اظهرت قدرة المستوى المتوسط من التمييز بين مجموعتي مرتفعي ومنخفضي المعدل التراكمي باعتباره مؤشر على المستوى الحقيقي للطالب.

كما تؤكد هذه النتائج دراسة Richard (2001) أن المفردات غير الموثوق بها هي المفردات التي تمتلك مستوى تمييز إما مرتفع أو منخفض بشدة، كما أشارت الدراسة أن الاختبار الذي يقيس مدى متسع من الصفة المقاسة ينبغي أن يتم اختيار مفراته بناء على مؤشر التمييز الذي يفيد بامتلاك المفردات لمستوى متوسط من التمييز.

**نتائج التحليل التمييزي لمستوى أقصى معلومات:**

أسفرت التحليلات التي اعتمدت على المستويات الثلاث من متغير أقصى معلومات (منخفض- متوسط- مرتفع) (كتغيرات ثلاثة مستقلة تمييزية) ومتغير مجموعتي مرتفعي ومنخفضي المعدل التراكمي (كتغير تابع تصنيفي) عن النتائج الموضحة بالجداول التالية:

## جدول (٢٠)

## الإحصاءات الوصفية لمتغيرات أقصى معلومات

## كما يبدو من نتائج التحليل التمييزي

| المجموعة | مستويات متغير المعلومات | المتوسط الحسابي | الإنحراف المعياري |
|---|---|---|---|
| منخفضي المعدل التراكمي | منخفض | ١٥.٨٨٣ | ٤.٦٧١٨٤ |
| | متوسط | ١٤.٩١١٣ | ٤.١٧٥٣٣ |
| | مرتفع | ١٤.١٢٠٦ | ٣.٧٧٠٠٤ |
| مرتفعي المعدل التراكمي | منخفض | ٢٣.٦٣٠١ | ٢.١٩٠٩٣ |
| | متوسط | ٢٣.٩٨٦٣ | ٢.٧٦٦٧٦ |
| | مرتفع | ٢١.٠٢٠٥ | ٢.٨٥٣٨٤ |

يتبين من النتائج الموضحة بجدول (٢٠) وجود فروق ظاهرية بين متوسطات درجات عينتي منخفضي ومرتفعي المعدل التراكمي على أنواع الأسئلة التي تختلف باختلاف مستوى أقصى معلومات (منخفض، متوسط، مرتفع)، بلغ أعلى متوسط القيمة (٢٣.٩٨٦٣) وكان لدرجات عينة مرتفعي المعدل التراكمي على الأسئلة التي تندرج تحت المستوى المتوسط لأقصى معلومات، بينما بلغ أدنى متوسط القيمة (١٤.١٢٠٦) وكان لعينة منخفضي المعدل التراكمي على الأسئلة التي تندرج تحت مستوى أقصى معلومات.

## جدول (٢١)

## جدول اختبار Box's M لدلالة تجانس تباين أو تغاير مجموعتي المرتفعين والمنخفضين

| | |
|---|---|
| اختبار Box's M | ٦.٩٤٥ |
| مستوى الدلالة | ٠.٣٦٦ |

ويؤكد الجدول (٢١)، أن قيمة (ف) لاختبار (Box's M) بلغت (٦.٩٤٥) وهي قيمة غير دالة، وبالتالي فإن البيانات تستوفي شرط الاعتدالية وتجانس التباين والتغاير.

**جدول (٢٢)**

**معلومات حول خطوات دخول المتغيرات ودلالة كل منها في التحليل التمييزي**

| الخطوة | | اختبار Tolerance | Sig of. F to Remove | Wilks' Lambda |
|---|---|---|---|---|
| ٢ | معلومات متوسطة | ٠.٥٥٥ | ٠.٠٠٠ | ٠.٥١٦ |
| | معلومات مرتفعة | ٠.٠٠٠ | ٠.٠٠٠ | ٠.٤٩٨ |

يتضح من الجدول (٢٢) ترتيب دخول المتغيرات الثلاثة في النموذج. وقد تم ترتيب هذه المتغيرات وفقاً لتأثير كل منها أو نسبة التباين بين المجموعتين الذي يفسره كل منها؛ حيث يتم البدء بالمتغير الأكثر تأثيراً وهو المستوى المتوسط من المعلومات ثم المستوى المرتفع من نفس المتغير.

**جدول (٢٣)**

**قيمة الجذر الكامن ومعامل الارتباط القانوني**

| الدالة | الجذر الكامن | نسبة التباين % | النسبة التراكمية للتباين % | معامل الإرتباط القانوني | مربع إيتا |
|---|---|---|---|---|---|
| ١ | a١.٣٩٢ | ١٠٠ | ١٠٠ | ٠.٧٨٣ | ٠.٦١ |

يتضح من الجدول (٢٣) أن قيمة الجذر الكامن بلغت (١.٣٩٢) كما أن قيمة الارتباط القانوني قد بلغت (٠.٧٨٣)، وهو يشير إلى قوة ارتباط الدرجات التمييزية وفئات المتغير النوعي أو المجموعات. وبتربيع قيمة هذا المعامل تم الحصول على قيمة مربع إيتا التي بلغت (٠.٦١)،

وهو أحد مؤشرات حجم الأثر التي تعتمد على تحديد نسبة التباين المفسر.

**جدول (٢٤)**

**قيمة معامل ويلكس لامدا ودلالته الإحصائية**

| اختبار الدوال | اختبار قيمة Wilks' Lambda | مربع كاي | درجات الحرية | مستوى الدلالة |
|---|---|---|---|---|
| ١ | ٠.٤١٨ | ٣٧٠.٧١٦ | ٢ | ٠.٠٠٠ |

ويبين الجدول (٢٤) أن قيمة معامل ويلكس لامبدا بلغت (٠.٤١٨)، وهي دالة عند مستوى (٠.٠١). ومعامل ويلكس لامبدا هو مؤشر على كمية التباين غير المفسر في الدرجات التمييزية، وكلما صغرت قيمة هذا المعامل كانت نتائج التحليل أفضل؛ حيث تدل القيمة الأقرب إلى الصفر على اختلاف المجموعات.

**جدول (٢٥)**

**المعاملات المعيارية للدالة التمييزية**

| المتغيرات التمييزية | المعاملات المعيارية |
|---|---|
| المستوى المتوسط للمعلومات | ٠.٧٦٧ |
| المستوى المرتفع للمعلومات | ٠.٣٠٩ |

ويتضح من الجدول (٢٥) الذي يوضح معاملات المعيارية للدالة التمييزية، أن القيم الواردة للمعاملات تبين أن مستوى دالة المعلومات المتوسط من حيث اقصى معلومات بلغ معامله المعياري القيمة (٠.٧٦٧) وهو اكبر من المعامل المعياري لمتغير المستوى الأعلى من أقصى معلومات والذي بلغت قيمته (٠.٣٠٩) وبالتالي يتضح أن المستوى المتوسط من متغير اقصى معلومات كان مؤثرا في النموذج

بدرجة تفوق تأثير المستوى المرتفع من نفس المتغير، وهذه المعاملات مفيدة في المقارنة بين أثر كل متغير كمي في النموذج.

**جدول (٢٦)**

**معامــلات الدالة التمييزيــة**

| المتغيرات التمييزية | معاملات الدالة التمييزية |
|---|---|
| المستوى المتوسط للمعلومات | ٠.٣٨٥ |
| المستوى المرتفع للمعلومات | ٠.٢٠٤ |
| الثابت | ٥.١١١- |

يتبين من الجدول (٢٦) معاملات الدالة التمييزية لكل متغير من المتغيرات التمييزية (المستقلة) وهذه القيم هي قيم (أ) للمستوى المتوسط من متغير أقصى معلومات بلغت قيمتها (٠.٣٨٥)، و(ب) لمتغير المستويات العليا من متغير أقصى معلومات وبلغت قيمته (٠.٢٠٤)، بالإضافة على ثابت المعادلة (ج) والذي بلغت قيمته (-٥.١١١) وهي القيم اللازمة لمعادلة الانحدار التي تمكننا من التنبؤ بالمجموعة التي تنتمي إليها الحالات الجديدة، كما سيرد لا حقاً.

**جدول (٢٧)**

**إقتران تصنيف عضوية أفراد فئتي منخفضي ومرتفعي التراكمي بالتصنيف المتنبا به من نموذج التحليل**

| المجموعة الأصلية المصنفة في ضوء المعدل التراكمي | المجموعة التنبأ بها من التحليل | | المجموع |
|---|---|---|---|
| | منخفضين | منخفضين | |
| منخفضين | ٢٦٤ | ١٨ | ٢٨٢ |
| مرتفعين | ١٨ | ٢٢٨ | ٢٤٦ |

يتضح من جدول (٢٧) أن ست وثلاثون حالة كان التنبؤ بها خاطئاً حيث تم تصنيفهم من قبل التحليل في مجموعات خاطئة فقد صنفت (١٨) حالات من مرتفعي المعدل التراكمي بفئة منخفضي المعدل التراكمي، بينما صنف نفس العدد أيضاً من حالات منخفضي المعدل التراكمي بفئة مرتفعي المعدل التراكمي، بينما بقية الحالات (٢٦٤ + ٢٢٨ = ٤٩٢) كان التنبؤ بها صحيحاً، وهذا يشير إلى أن نسبة صحة التنبؤ تبلغ (٤٩٢ ÷ ٥٢٨ = ٩٣.٢%) تقريباً.

**جدول (٢٨)**

**معامل قوة الاتفاق لكابا بين التصنيف المتنبأ به والتصنيف الأصلي لأفراد فئتي منخفضي ومرتفعي المعدل التراكمي**

| اختبار Kappa | القيمة | مستوى الدلالة |
|---|---|---|
| | ٠.٨٢٢ | ٠ |

وتؤكد قيمة معامل كابا من جدول (٢٨) على فاعلية النموذج في التنبؤ، فقد بلغت (٠.٨٢) مما يدل على وجود ارتباط قوي بين النتيجة الفعلية والنتيجة المتنبأ بها. مما يشير إلى قوة اتساق التصنيف المتوقع من النموذج الإحصائي إعتماد على المتغيرات التمييزية المتضمنة به (المستوى المتوسط من أقصى معلومات والمستوى المرتفع من اقصى معلومات) لأفراد عينة البحث مع تصنيفهم الأصلي بمجموعتي منخفضي ومرتفعي المعدل التراكمي.

يتضح مما سبق من نتائج أن الأسئلة الاختبارية التي تمتلك قدرا متوسطاً أو مرتفعا من متغير أقصى معلومات كانتا أكثر تأثيرا في نموذج التحليل وبالتالي كانتا أكثر قدرة على التمييز بين عينتي مرتفعي

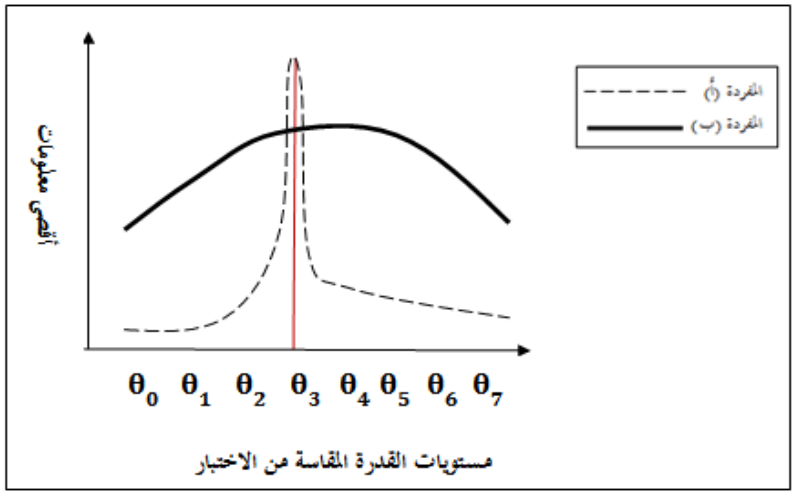
ومنخفضي المعدل التراكمي، بما يفيد بأن هذين المستويين يمكن الاعتماد عليهما في تقييم كفاءة الاختبار التحصيلي البنكي، وهذه النتائج أكدتها الأطر النظرية والأدبيات التي أفادت بأن المفردة الاختبارية التي تمتلك أعلى قيمة للمعلومات عند مستوى معين من مستويات القدرة او الصفة المستهدف قياسها من نفس الاختبار تكون أفضل من مفردات أخرى تمتلك قدراً أقل من المعلومات عند نفس المستوى من القدرة (Edward 2010: 474)، كما أثبتت العديد من الدراسة والادبيات أن اختيار المفردات الاختبارية التي تمتلك المستوى الأعلى من متغير اقصى معلومات تعطي تقديرات ادق للقدرة المقاسة Aslam & MuhammadAltaf, Saim (2011) ,، بما يعني أن مؤشر أقصى معلومات تمتلكه المفردات الاختبارية يمكن أين يكون بين المؤشرات التي يمكن الاعتماد عليها في تحديد الكفاءة النسبية للاختبار، وقد أكدت هذه النتيجة العديد من الأدبيات التي أثبتت أن تمتع اختبارات الموائمة المحوسبة بالعديد من المزايا من أهمها إمكانية تقدير قدرة المفحوص باستخدام عدد قليل من المفردات التي تتميز بامتلاكها لأقصى معلومات (Siang & Fritz,2006,241)، ومن المعروف أن من بين الطرق المستخدمة في اختيار المفردة في هذه النوعية من اختبارات الموائمة هي طريقة أقصى معلومات(Schmitt, 2010:233)، ولذلك يؤكد Marinagi, & Kaburlasos (2010) على أن أسلوب المعلومات القصوى يعطي نتائج جيدة في اختيار مفردات الاختبار ويؤدي إلى تحقيق مزيد من الدقة في قياس القدرة المستهدفة بالاختبار إذا قورن بأساليب أخرى في اختيار مفردات الاختبار، وقد أثبتت أيضاً دراسة Aslam & MuhammadAltaf, Saim (2011) , أن أسلوب

أقصى معلومات تفوق على أسلوب أقصى تمايز كأسلوبين لانتقاء المفردة وذلك في الاختبارات الخطية والتواؤمية معا، وهذا ما يدعم أيضاً النتائج التي توصلت إليها الدراسة الحالية، كما أن ما يؤكد هذه النتيجة أيضا هو ما أشار إليه Kingsbury & Zara (2002) بأن هناك العديد من المميزات التي يتمتع بها أسلوب أقصى معلومات في انتقاء المفردات الاختبارية حيث أنه يوفر أعلى كفاءة للاختبار، كما أن انتقاء المفردة بأسلوب المعلومات القصوى.

كما تشير دراسة Lihua Yao (2011) يعطي أقل خطأ قياس عند مستوى القدرة المستهدفة، وقد جاء ت النتائج الخاصة بتفوق قدرة الأسئلة التي تمتلك قدر متوسط من أقصى معلومات في التمييز بين مجموعتي مرتفعي ومنخفضي المعدل التراكمي عن قدرة الأسئلة التي تمتلك قدر مرتفع من أقصى معلومات بصورة ليست متوقعة، وقد تفسر هذه النتائج بأن بعض الأسئلة قد تمتلك قدر مرتفع بشكل متطرف للمعلومان عند نقطة محددة من القدرة المقاسة وقد تمتلك نفس المفردة قيم منخفضة للمعلومات عند نقاط أخرى على متصل القدرة وبالتالي فإن هذه المفردة تصبح أقل تمييزا بين الأفراد عند جيع مستويات القدرة المقاسة، وذلك إذا ماقورنت بالمفردة التي تمتلك قدرا متوسطا من المعلومات عند غالبية نقاط القدرة المقاسة وبالتالي فتكون المفردة الأخيرة هي الأكثر تمييزا بين الأفراد.

والشكل التالي يوضح ذلك حيث تمتلك المفردة (أ) قيمة مرتفعة جدا من متغير أقصى معلومات عند نقطة محددة (مستوى محدد)من القدرة وهي $\theta_3$ بينما تمتلك قيم منخفضة نفس المتغير عند نقاط القدرة المتبقية،

في حين أن المفردة (ب) تمتلك مستوى متوسط من المعلومات عند مختلف نقاط (مستويات) القدرة من $\theta_0$ حتى $\theta_7$ وبالتالي تكون ادق وأفضل من المفردة (أ) التي لاتمتلك إلا معلومات قصوى فقط عند نقطة وحيدة من نقاط القدرة ومعلومات منخفضة عند مختلف نقاط القدرة الأخرى.




**شكل (١)**

**مقارنة بين مفردتان تمتلك إحداها مستوى مرتفع من المعلومات قصوى عند نقطة محددة والأخرى تمتلك مستوى متوسط من المعلومات على اغلب نقاط القدرة المقاسة**

## نتائج التحليل التمييزي لنتائج متغير مستوى التخمين:

أسفرت التحليلات التي اعتمدت على المستويات الثلاثة من متغير مستوى التخمين (منخفض- متوسط- مرتفع) (كتغيرات ثلاثة مستقلة تمييزية)، ومتغير مجموعتي مرتفعي ومنخفضي المعدل التراكمي (كتغير تابع تصنيفي) عن النتائج الموضحة بالجداول التالية:

**جدول (٢٩)**

**جدول الإحصاءات الوصفية لمتغيرات مستوى التخمين كما يبدو من نتائج التحليل التمييزي**

| المجموعة | مستويات متغير التخمين | المتوسط الحسابي | الإنحراف المعياري |
|---|---|---|---|
| منخفضي المعدل التراكمي | منخفض | ١٢.٩٧٥٢ | ٣.٧٧٨٩٦ |
| | متوسط | ١٢.١٤٥٤ | ٣.٥١٢٥٨ |
| | مرتفع | ١٣.٥٩٥٧ | ٣.٥٤٢٢٨ |
| مرتفعي المعدل التراكمي | منخفض | ٢٠.٢٦٧١ | ٢.٥٨٩٨٨ |
| | متوسط | ١٧.٩١٧٨ | ١.٤٩٧١٦ |
| | مرتفع | ٢٠.٣١٥١ | ٣.٠٤١٦٧ |

يتبين من النتائج الموضحة بالجدول (٢٩) وجود فروق ظاهرية بين متوسطات درجات عينتي منخفضي ومرتفعي المعدل التراكمي على أنواع الأسئلة التي تختلف باختلاف مستوى التخمين (منخفضة، متوسطة، مرتفعة)، بلغ أعلى متوسط القيمة (٢٠.٣١٥١)، وكان لدرجات عينة مرتفعي المعدل التراكمي على الأسئلة التي تمتلك مستوى مرتفع من التخمين، بينما بلغ أدنى متوسط القيمة (١٢.١٤٥٤) وكان لعينة منخفضي المعدل التراكمي على الأسئلة التي تمتلك مستوى متوسط من التخمين.

**جدول (٣٠)**

**جدول اختبار Box's M لدلالة تجانس تباين أو تغاير مجموعتي المرتفعين والمنخفضين بالتحليلات الخاصة بمتغيرات التخمين**

| | |
|---|---|
| اختبار Box's M | ١.٨٢٤ |
| مستوى الدلالة | ٠.٠٩ |

ويؤكد الجدول (٣٠)، أن قيمة (ف) لاختبار (Box's M) بلغت (١.٨٢٤) وهي قيمة غير دالة، وبالتالي فإن البيانات تستوفي شرط الاعتدالية وتجانس التباين والتغاير.

**جدول (٣١)**

**معلومات حول خطوات دخول المتغيرات ودلالة كل منها في التحليل التمييزي لمتغيرات مستوى التخمين**

| الخطوة | | Tolerance | Sig of F. to Remove | Wilks' Lambda |
|---|---|---|---|---|
| ٣ | مستوى التخمين المنخفض | ٠.٣٤ | ٠.٠٠٠ | ٠.٤٨٤ |
| | مستوى التخمين المتوسط | ٠.٤٩٥ | ٠.٠٠٠ | ٠.٤٧٨ |
| | مستوى التخمين المرتفع | ٠.٣٩١ | ٠.٠٠٥ | ٠.٤٧٢ |

يتضح من الجدول (٣١) من الخطوة الأخيرة رقم (٣) ترتيب دخول المتغيرات الثلاثة في النموذج. وقد تم ترتيب هذه المتغيرات وفقاً لتأثير كل منها أو نسبة التباين بين المجموعتين الذي يفسره كل منها؛ حيث يتم البدء بالمتغير الأكثر تأثيراً وهو مستوى التخمين المنخفض ثم المستوى المتوسط من نفس المتغير فالمستوى المرتفع.

**جدول (٣٢)**

**قيمة الجذر الكامن ومعامل الارتباط القانوني للتحليلات الخاصة بمستويات متغير التخمين**

| الدالة | الجذر الكامن | نسبة التباين % | النسبة التراكمية للتباين % | معامل الإرتباط القانوني | مربع إيتا |
|---|---|---|---|---|---|
| ١ | a١.١٥٧ | ١٠٠ | ١٠٠ | ٠.٧٣٢ | ٠.٥٣٦ |

يتضح من الجدول (٣٢) أن قيمة الجذر الكامن بلغت (١.١٥٧) كما أن قيمة الارتباط القانوني قد بلغت (٠.٧٣٢)، وهو يشير إلى مدى الارتباط بين الدرجات التمييزية وفئات المتغير النوعي أو المجموعات.

وبتربيع قيمة هذا المعامل تم الحصول على قيمة مربع إيتا التي بلغت (٠.٥٤)، تقريبا وهو أحد مؤشرات حجم الأثر التي تعتمد على تحديد نسبة التباين المفسر.

**جدول (٣٣)**

**قيمة معامل ويلكس لامدا ودلالته الإحصائية**

| اختبار الدوال | اختبار قيمة Wilks' Lambda | مربع كاي | درجات الحرية | مستوى الدلالة |
|---|---|---|---|---|
| ١ | ٠.٤٦٤ | ٣٢٦.٢٨٩ | ٣ | ٠.٠٠٠ |

ويبين الجدول (٣٣) أن قيمة معامل ويلكس لامبدا بلغت (٠.٤٦٤)، وهي دالة عند مستوى (٠.٠١). ومعامل ويلكس لامبدا هو مؤشر على كمية التباين غير المفسر في الدرجات التمييزية، وكلما صغرت قيمة هذا المعامل كانت نتائج التحليل أفضل؛ حيث تدل القيمة الأقرب إلى الصفر على اختلاف المجموعات.

**جدول (٣٤)**

**المعامــلات المعيارية للدالة التمييزيــة**

| المتغيرات التمييزية | المعاملات المعيارية |
|---|---|
| مستوى التخمين المنخفض | ٠.٤٨٥ |
| مستوى التخمين المتوسط | ٠.٣٣١ |
| مستوى التخمين المرتفع | ٠.٢٩٧ |

ويتضح من الجدول السابق (٣٤) الذي يوضح معاملات المعيارية للدالة التمييزية، أن القيم الواردة للمعاملات تبين أن المستوى المنخفض من متغير التخمين بلغ معامله المعياري القيمة (٠.٤٨٥) وهو أكبر من المعامل المعياري لمتغير المستوى المتوسط من متغير التخمين والذي بلغت قيمته (٠.٣٣١)، وكذلك المعامل المعياري للمستوى المرتفع من

متغير قوة تمييز المفردات والذي بلغت قيمته (٠.٢٩٧) وبالتالي يتضح أن المستوى المنخفض من متغير التخمين كان مؤثرا في النموذج بدرجة تفوق تأثير المستوى المتوسط وكذلك المنخفض من نفس المتغير.

**جدول (٣٥)**

**معامـلات الدالة التمييزيـة**

| المتغيرات التمييزية | المعاملات المعيارية |
|---|---|
| مستوى التخمين المنخفض | ٠.١٤٢ |
| مستوى التخمين المتوسط | ٠.١١١ |
| مستوى التخمين المرتفع | ٠.٠٨٨ |
| الثابت | ٥.١٥١- |

يتبين من الجدول (٣٥) معاملات الدالة التمييزية لكل متغير من المتغيرات التمييزية (المستقلة) وهذه القيم هي قيم (أ) للمستوى المنخفض من متغير التخمين بلغت قيمتها (٠.١٤٢)، و(ب) للمستوى المتوسط من متغير التخمين وبلغت قيمته (٠.١١١)، و(ج) للمستوى الرتفع من نفس المتغير وبلغت قيمته (٠.٠٨٨) بالإضافة على ثابت المعادلة (ث) والذي بلغت قيمته (-٥.١٥١) وهي القيم اللازمة لمعادلة الانحدار التي تمكننا من التنبؤ بالمجموعة التي تنتمي إليها الحالات الجديدة، كما سيرد لا حقاً.

**جدول (٣٦)**

**إقتران تصنيف عضوية أفراد فئتي منخفضي ومرتفعي التراكمي بالتصنيف المتنبا به من نموذج التحليل**

| المجموعة الأصلية المصنفة في ضوء المعدل التراكمي | المجموعة التنبأ بها من التحليل | | المجموع |
|---|---|---|---|
| | منخفضين | منخفضين | |
| منخفضين | ٢٥٦ | ٢٦ | ٢٨٢ |
| مرتفعين | ١٨ | ٢٢٨ | ٢٤٦ |

من خلال الجدول (٣٦) يتضح أن إثنى وعشرون حالة فقط (٢٦ + ١٨ = ٤٤) كان التنبؤ بها خاطئاً حيث تم تصنيفهم من قبل التحليل في مجموعات خاطئة فقد صنفت (١٨) حالات من مرتفعي المعدل التراكمي بفئة منخفضي المعدل التراكمي، بينما صنفت (٢٦) حالات من منخفضي المعدل التراكمي بفئة مرتفعي المعدل الراكمي، بينما بقية الحالات (٢٥٦ + ٢٢٨ = ٤٨٤) كان التنبؤ بها صحيحاً، وهذا يشير إلى أن نسبة صحة التنبؤ تبلغ (٤٨٤ ÷ ٥٢٨) = ٩١.٧% تقريباً.

**جدول (٣٧)**

**معامل قوة الاتفاق لكابا بين التصنيف المتنبأ والتصنيف الأصلي لأفراد فئتي منخفضي ومرتفعي المعدل التراكمي**

| اختبار Kappa | القيمة | مستوى الدالاة |
|---|---|---|
| | ٠.٨١٢ | ٠.٠٠٠ |

وتؤكد قيمة معامل كابا من جدول (٣٧) على فاعلية النموذج في التنبؤ، فقد بلغت (٠.٨١) تقريبا مما يدل على وجود ارتباط قوي بين النتيجة الفعلية والنتيجة المتنبأ بها. مما يشير إلى قوة اتساق التصنيف المتوقع من النموذج الإحصائي إعتماد على المتغيرات التمييزية المتضمنة به والممثلة بمستويات متغير التخمين الثلاث (منخفض، متوسط، مرتفع) لأفراد عينة البحث مع تصنيفهم الأصلي بمجموعتي منخفضي ومرتفعي المعدل التراكمي.

يتضح مما سبق من نتائج أن الأسئلة الاختبارية التي امتلكت مستويات منخفضة من التخمين تفوقت على الأسئلة الأخرى التي امتلكت قدراً مرتفعاً ومتوسطاً من نفس المتغير حيث كان للأسئلة منخفضة

التخمين الأثر الأكبر في نموذج التحليل وبالتالي كانت أكثر قدرة على التمييز بين عينتي مرتفعي ومنخفضي المعدل التراكمي، يلي ذلك الأسئلة متوسطة التخمين والأخرى ثم المرتفعة بما يفيد بأن الأسئلة التي تمتلك المستوى المنخفض من التخمين يمكن الاعتماد عليها في تقييم الكفاءة النسبية للاختبار ويمكن أن يكون لها وزنا أعلى في تقييم الكفاءة النسبية للاختبار.

وقد تاكدت هذه النتيجة من خلال بعض الأدبيات التي أشارت إلى أن سلوك التخمين يؤثر على تقديرات القدرة للمختبرين، فقد اظهرت دراسة Plale. Chamge, & Gene (2011) أن نسبة حدوث سلوك التخمين للمفردات من نوع الصواب والخطا جاءت اعلى بصورة جوهرية عن نظائرها التي بدت من خلال الإجابة عن أسئلة الاختيار من المتعدد وان هذه النوعية من الاستجابات غير الملائمة أثرت سلبا على دقة تقديرات القدرة للأفراد المتلقين للاختبار.

كما تؤكد هذه النتيجة مع ما أشارت إليه بعض الدراسات من التأثيرات التي تحدثها الاستجابات الناتجة عن التخمين على بعض خصائص الاختبار الإحصائية وتقدير الدرجات الحقيقية للمختبرين أظهرت دراسة Richard (2001) أن نسبة التخمين المرتفعة تؤثر سلبيا في تقدير قوة التمييز للمفردات وكذلك ثبات الاختبار ومن ثم كفاءة هذه الاختبارات في تقدير الدرجات الحقيقية للمختبرين.

وقد أكدت هذه النتيجة أيضاً النتائج التي توصلت إليها بعض الدراسات التي اهتمت بالتعرف على التغير الذي يحدثه التخمين على اداء الطالب المتصل بأنواع مختلفة من الفقرات الاختبارية، حيث

اظهرت دراسة Schurmeier, et al (2011) تأثير التغير في نوع الأسئلة على تغير اداء الممتحن، حيث تأثر اداء الطلاب يتاثر بشكل دال بين صياغة الأسئلة المصاغة باسلوب الصواب والخطأ التيي تمتلك مستويات اعلى من التخمين عن أسئلة الاختيار من متعدد، أو إذا صيغت بصورة تكملة أو مزاوجة وهي انواع المفردات، وكذلك إذا استخدمت صيغت بأسلوب المقال وهي التي تمتلك بمستويات أقل من التخمين.

ويتضح من خلال المناقشة السابقة أن المستوى المنخفض من التخمين يؤدي إلى تقديرات حقيقية لمستويات اداء المختبرين ومن ثم يمكن الاستناد إليه كمعيار لتقييم الكفاءة النسبية للاختبار

**ثانياً: نتائج الإجابة عن السؤال الثاني من أسئلة الدراسة:**

أسفرت إجراءات الإجابة عن السؤال الثاني والذي ينص على: ما دلالة القوة التمييزية لكل متغير من متغيرات (نوع المفردات، المستوى المعرفي للمفردات، تمثيل المفردات نطاق السلوك المقاس) بين مرتفعي ومنخفضي المعدل التراكمي التحصيلي؟ عن النتائج التالية:

*** نتائج التحليل التمييزي لمتغير نوع المفردة:**

أسفرت التحليلات التي اعتمدت على المستويات السنة من لمتغير نوع المفردة (اختيار من متعدد- صواب وخطأ- مزاوجة- إكمال- إجابات قصيرة- مقال) (كتغيرات مستقلة تمييزية) ومتغير مجموعتي مرتفعي ومنخفضي المعدل التراكمي (كتغير تابع تصنيفي) عن النتائج الموضحة بالجداول التالية:

## جدول (٣٨)

## جدول الإحصاءات الوصفية لمتغيرات مستوى التخمين كما يبدو من نتائج التحليل التمييزي

| المجموعة | متغير نوع المفردة | المجموعة | مستويات متغير المعلومات |
|---|---|---|---|
| منخفضي المعدل التراكمي | اختيار من متعدد | ٩.٦٨ | ٢.٧٦٨ |
| | صواب وخطأ | ١٢.٨٦ | ٢.٦٣٥ |
| | مزاوجة | ٨.٣٩ | ٤.٨٤ |
| | إكمال | ٦.٢٦ | ٣.٧٧٦ |
| | إجابة قصيرة | ٣.٥١ | ٢.١٤٥ |
| | مقال | ٠.٦٩ | ١.٢٧٣ |
| مرتفعي المعدل التراكمي | اختيار من متعدد | ١٥.٥٥ | ١.٣٨ |
| | صواب وخطأ | ١٦.٩٩ | ١.٦٤٨ |
| | مزاوجة | ١٨.١٩ | ١.٥٥٩ |
| | إكمال | ١٤.٦٤ | ٢.٢٧٧ |
| | إجابة قصيرة | ٥.٥ | ٠.٩٧٧ |
| | مقال | ٤.٠٧ | ١.٣٥٨ |

يتبين من النتائج الموضحة بالجدول (٣٨) وجود فروق ظاهرية بين متوسطات درجات عينتي منخفضي ومرتفعي المعدل التراكمي على أسئلة الاختبار التحصيلي التي تختلف باختلاف النوع (اختيار من متعدد، صواب وخطأ، المزاوجة، الإكمال، الأسئلة القصيرة، المقال)، حيث بلغ أعلى متوسط القيمة (١٨.١٩)، وكان لدرجات عينة مرتفعي المعدل التراكمي على أسئلة المزاوجة، بينما بلغ أدنى متوسط القيمة (٣.٥١) وكان لعينة منخفضي المعدل التراكمي على أسئلة الإجابة القصيرة.

**جدول (٣٩)**

**دلالة اختبار Box's M لتجانس تباين أو تغاير مجموعتي المرتفعين والمنخفضين بالتحليلات الخاصة بمتغير نوع المفردة**

| | |
|---|---|
| اختبار Box's M | ٢.٨١٤ |
| مستوى الدلالة | ٠.١٢ |

ويؤكد الجدول (٣٩)، أن قيمة (ف) لاختبار (Box's M) بلغت (٢.٨١٤) وهي قيمة غير دالة، وبالتالي فإن البيانات تستوفي شرط الاعتدالية وتجانس التباين والتغاير.

**جدول (٤١)**

**معلومات حول خطوات دخول المتغيرات ودلالة كل منها في التحليل التمييزي لمتغيرات نوع المفردات**

| | الخطوة | Tolerance | Sig. of F to Remove | Wilks' Lambda |
|---|---|---|---|---|
| ٧ | مقال | ٠.٧٩٣ | ٠.٠٠٠ | ٠.٢٥٧ |
| | اختيار من متعدد | ٠.٧٧٧ | ٠.٠٠٠ | ٠.٢٢٨ |
| | مزاوجة | ٠.٧٤٩ | ٠.٠٠٠ | ٠.٢٥٦ |
| | صواب وخطأ | ٠.٧٧٦ | ٠.٠٠٧ | ٠.٢١٥ |
| | إجابة قصيرة | ٠.٦٢٤ | ٠.٠١٥ | ٠.٢١٤ |

يتضح من الجدول (٤٠) من آخرخطوة (٧) عدم تجاوز أي متغير النسبة (٠.١٠) المحددة لدلالة الخروج من نموذج التحليل Sig. of F to Remove، ورتبت المتغيرات وفقاً لقيمة اختبار "ويلكس لامدا" لنسبة التباين بين المجموعتين الذي يفسره كل منها؛ حيث يتم البدء بالمتغير الأكثر تأثيراً ثم المستوى المتوسط ثم المستوى المنخفض.

**جدول (٤١)**

**قيمة الجذر الكامن ومعامل الارتباط القانوني للتحليلات الخاصة بمستويات متغير التخمين**

| الدالة | الجذر الكامن | نسبة التباين % | النسبة التراكمية للتباين % | معامل الإرتباط القانوني |
|---|---|---|---|---|
| ١ | a٣.٧٥٣ | ١٠٠ | ١٠٠ | ٠.٨٨٩ |

يتضح من الجدول (٤١) أن قيمة الجذر الكامن بلغت (٣.٧٥٣) كما أن قيمة الارتباط القانوني قد بلغت (٠.٨٨٩)، وهو يشير إلى مدى الارتباط بين الدرجات التمييزية وفئات المتغير النوعي أو المجموعات.

وبتربيع قيمة هذا المعامل تم الحصول على قيمة مربع إيتا التي بلغت (٠.٧٩)، تقريبا وهو أحد مؤشرات حجم الأثر التي تعتمد على تحديد نسبة التباين المفسر.

**جدول (٤٢)**

**قيمة معامل ويلكس لامدا ودلالته الإحصائية**

| اختبار الدوال | اختبار قيمة Wilks' Lambda | مربع كاي | درجات الحرية | مستوى الدلالة |
|---|---|---|---|---|
| ١ | ٠.٢١ | ٥٠٥.٨٠٩ | ٥ | ٠ |

ويبين الجدول (٤٢) أن قيمة معامل ويلكس لامبدا بلغت (٠.٢١)، وهي دالة عند مستوى (٠.٠١).

ومعامل ويلكس لامبدا هو مؤشر على كمية التباين غير المفسر في الدرجات التمييزية، وكلما صغرت قيمة هذا المعامل كانت نتائج التحليل أفضل؛ حيث تدل القيمة الأقرب إلى الصفر على اختلاف المجموعات.

**جدول (٤٣)**

**المعامـلات المعيارية للدالة التمييزيـة**

| المتغيرات التمييزية | المعاملات المعيارية |
|---|---|
| اختيار من متعدد | ٠.٣٥٢ |
| صواب وخطأ | ٠.١٩٢ |
| مزاوجة | ٠.٥٣٩ |
| إجابة قصيرة | ١٠.١٩١ |
| مقال | ٠.٥٥١ |

ويتضح من الجدول السابق (٤٤) الذي يوضح المعاملات المعيارية للدالة التمييزية.

أن القيم الواردة للمعاملات تبين أن المفردات من نوع المقال بلغ معامله المعياري القيمة (٠.٥٥١) وهو أكبر من المعامل المعياري للمفردات من نوع المزاوجة والذي بلغت قيمته (٠.٥٣٩)، وكذلك المعامل المعياري للمفردة من نوع الاختيار من متعدد والذي بلغت قيمته (٠.٣٥٢) أكبر من المعامل المعياري لأسئلة الصواب والخطأ الذي بلغت قيمته (٠.١٩٢).

هذا، وقد جاء المعامل المعياري للأسئلة، من نوع الإجابة القصيرة في المرتبة الخيرة حيث بلغت قيمة المعامل المعياري لها (-٠.١٩١).

وبالتالي فإن تاثير الأسئلة من نوع المزاوجة كانت مؤثرة في النموذج بدرجة تفوق تأثير أنواع الأسئلة الأخرى.

## جدول (٤٥)

## معامـلات الدالة التمييزيـة

| المتغيرات التمييزية | المعاملات المعيارية |
|---|---|
| اختيار من متعدد | ٠.١٥٦ |
| صواب وخطأ | ٠.٠٨٥ |
| مزاوجة | ٠.١٤٧ |
| إجابة قصيرة | -٠.١١١ |
| اختيار من متعدد | ٠.٤١١ |
| الثابت | -٥.٤٤٣ |

يتبين من الجدول (٤٥) معاملات الدالة التمييزية لكل متغير من المتغيرات التمييزية (المستقلة) وهذه القيم هي قيم (أ) للأسئلة من نوع الاختيار من متعدد بلغت قيمتها (٠.١٥٦)، و(ب) للأسئلة من نوع الصواب والخطأ وبلغت قيمته (٠.٠٨٥)، و(ج) لأسئلة المزاوجة وبلغت قيمته (٠.١٤٧)، (د) لأسئلة الإجابة القصيرة وبلغت قيمته (٠.١١١)، (هـ) لأسئلة المقال وبلغت قيمته (٠.٤١١) بالإضافة على ثابت المعادلة (ث) والذي بلغت قيمته (-٥.٤٤٣) وهي القيم اللازمة لمعادلة الانحدار التي تمكننا من التنبؤ بالمجموعة التي تنتمي إليها الحالات الجديدة.

## جدول (٤٦)

## إقتران تصنيف عضوية أفراد فئتي منخفضي ومرتفعي التراكمي بالتصنيف المتنبا به من نموذج التحليل

| المجموعة الأصلية المصنفة في ضوء المعدل التراكمي | المجموعة التنبأ بها من التحليل | | المجموع |
|---|---|---|---|
| | منخفضين | منخفضين | |
| منخفضين | ٢٦٣ | ١٩ | ٢٨٢ |
| مرتفعين | ١١ | ٢٣٥ | ٢٤٦ |

يتضح من جدول (٤٦) أن إثنى وعشرون حالة فقط (١٩ + ١١ = ٣٠) كان التنبؤ بها خاطئاً حيث تم تصنيفهم من قبل التحليل في مجموعات خاطئة فقد صنفت (١١) حالات من مرتفعي المعدل التراكمي بفئة منخفضي المعدل التراكمي، بينما صنفت (١٩) حالات من منخفضي المعدل التراكمي بفئة مرتفعي المعدل الراكمي، بينما بقية الحالات (٢٦٣ + ١٣٥ = ٣٩٨) كان التنبؤ بها صحيحاً، وهذا يشير إلى أن نسبة صحة التنبؤ تبلغ (٣٩٨ ÷ ٥٢٨)= ٩٤% تقريباً.

**جدول (٤٧)**

**معامل قوة الاتفاق لكابا بين التصنيف المتنبأ به والتصنيف الأصلي لأفراد فئتي منخفضي ومرتفعي المعدل التراكمي**

| اختبار Kappa | القيمة | مستوى الدالاة |
|---|---|---|
| | ٠.٩٥٧ | ٠ |

وتؤكد قيمة معامل كابا من جدول (٤٧)على فاعلية النموذج في التنبؤ، فقد بلغت (٠.٩٦) تقريبا مما يدل على وجود ارتباط قوي بين النتيجة الفعلية والنتيجة المتنبأ بها، مما يشير إلى قوة اتساق التصنيف المتوقع من النموذج الإحصائي إعتماداً على المتغيرات التمييزية المتضمنة به الممثلة بنوع السؤال لأفراد عينة البحث مع تصنيفهم بمجموعتي منخفضي ومرتفعي المعدل التراكمي.

يتضح مما سبق من نتائج أن الأسئلة من نوع المقال تفوقت على الأسئلة الأخرى، بينمالم تبدو أسئلة الإكمال أية تأثيرات مما أدى إلى عدم ظهورها بنموذج التحليل، حيث كان للأسئلة المقالية الأثر الأكبر في نموذج التحليل وبالتالي كانت أكثر قدرة على التمييز بين مرتفعي ومنخفضي المعدل التراكمي، وقد أكدت النتائج الخاصة بتفوق قدرة أسئلة

المقال على التمييز بين عينتي الدراسة من مرتفعي ومنخفضي المعدل التراكمي على غيرها من انواع الأسئلة الأخرى ما أكدته دراسة Hicks, et al (2010) التي أثبتت أن أسئلة (المقال) والتي تهتم بالنواحي التطبيقية هي أفضل الأسئلة التي تستخدم في الكشف عن الخبرات التعليمية التي اكتسبها الطلاب في بالبرامج التعليمية. نظرا لأن الأسئلة المقالية تمثل مجال واسع من القدرات والمهارات المعرفية (علام، ٢٠١٠: ١١٤-١١٥) كما أنها تقيس القدرة على تنظيم الأفكار والمعلومات (نصار، ٢٠١٠: ٢٢٣)، أما تفوق أسئلة المزاوجة على أسئلة الاختيار من متعدد وأسئلة الصواب والخطأ ثم في النهاية أسئلة الإجابة القصيرة قد تفسر بعدم تعلقها الشديد بعامل التخمين، نظراً لأنها تحتوي غالبا على عدد من بدائل الإجابة يفوق عدد البدائل الخاصة بأسئلة الاختيار من متعدد مما يقلص فرصة التخمين للإجابة على هذه النوعية من الأسئلة، ثم تاتي أسئلة الصواب والخطأ التي تمتلك أعلى مستويات التخمين (نصار، ٢٠١٠: ١٤٤) في التريب بما يفيد بأن أنواع الأسئلة التي تمتلك المستوى المنخفض من التخمين يمكن الاعتماد عليها في تقييم الكفاءة النسبية للاختبار ويمكن أن يكون لها وزنا أعلى في تقييم الكفاءة النسبية للاختبار، أما فيما يتعلق بالأسئلة الخاصة بالإجابة القصيرة فقد اتضح تأثيراتها الضئيلة بنموذج التحليل وقد يفسر ذلك بأن هذه النوعية قد يكون لها فعالية في قياس بالمستويات العقلية الدنيا أكثر من فعاليتها في قياس المستويات العليا، وتتفق تلك النتائج مع ما أكدته بعض الأدبيات بانتشار وشيوع استخدام بعض أنواع الأسئلة التي أثبتت الدراسة الحالية فعاليتها التميزية بين مرتفعي ومنخفضي التحصيل الدراسي.

حيث أظهرت دراسة Burton (2005) التي اشارت غالبية الجامعات الأمريكية تعتمد بصورة أساسية على نوعي أسئلة الاختيار من متعدد والصواب والخطأ بهذه الاختبارات، كما أكدت بعض الأدبيات تأثر أداء الممتحن وثبات الاختبار بنوع أسئلة الاختبار وبصفة خاصة الأسئلة التي أثبتت الدراسة الحالية قدرتها التنبؤية والتميزية بين مر تفعي ومنخفضي التحصيل، فقد أظهرت دراسة مصطفى (٢٠٠٠) كفاءة أسئلة الاختيار من متعدد إذا ماقونت بأسئلة الصواب والخطأ، كما أن أنواع أسئلة الاختبار وبصفة خاصة نوعي أسئلة الاختيار من متعدد والصواب والخطأ يمكن أن تؤثر في بعض الخصائص الإحصائية للاختبار وبالتالي على كفاءة الاختبار وهذا ما أكدته دراسة Richard (2001)، وقد أكدت أيضا بعض الدراسات قدرة أسئلة الاختيار متعدد على رفع ثبات الاختبار وبالتالي رفع كفاءة الاختبار فقد أظهرت دراسة Richard (2004) أن اسئلة الاختيار من متعدد لها تأثيرات إيجابية على ثبات الاختبار، وهي نفس النتائج الخاصة بأسئلة المزاوجة، كما أن الثبات يزداد طرديا مع زيادة عدد خيارات أو بدائل الإجابة المرتبطة بكل مفردة اختبارية، كما تؤكد نتائج الدراسة الحالية دراسة Jandaghi,& Shateria (2008) التي أظهرت أن نوعي أسئلة الاختيار من متعدد والصواب والخطأ لهما نفس القوة التنبؤية الخاصة بصدق الاختبار، كما أن اسئلة الاختيار من متعدد تؤدي إلى قيم أفضل نسبياً للثبات، وهذا يوحي بتفوق أسئلة الاختيار من متعدد نسبيا على أسئلة الصواب والخطأ وها ما يتفق مع نتائج الدراسة الحالية، كما أكدت دراسة Kucuk & Walters (2009) التي أوضحت أن نوعية المفردات التي تبتعد إلى حد ما عن التخمين مثل اسئلة الاختبار من متعدد وكذلك أسئلة المزاوجة

يمكن أن تشكل مؤشرا جيداً لكفاءة الاختبار، وقد تأكدت نفس النتائج مع النتائج التي أظهرتها دراسة Schurmeier, et al. (2010) التي أثبتت أن اداء الطلاب يتأثر بشكل دال بين صياغة الأسئلة باسلوب الصواب والخطا عن أسئلة الاختيار من متعدد، أو بصورة تكملة أو مزاوجة، وكذلك إذا صيغت بأسلوب المقال عن الأنواع الأخرى وقد تبين أن اسئلة المقال هي الأفضل أنواع الأسئلة في تقدير الأداء الحقيقي نظرا لابتعادها عن عامل التخمين، وقد أكدت بعض الدراسات فعالية اسئلة الاختيار من متعدد وكذلك المزاوجة وأيضا أسئلة المقال في الاختبارات التحصلية وفي قياس بعض بعض مهارات التفكير والمستويات المعرفية العليا، حيث أكدت دراسة Schriner Clifton (2010) & كفاءة أسئلة الاختيار من متعدد في قياس التفكير الناقد كما أن هذه النوعية لها القدرة في قياس المستويات المعرفية العليا ولكنها ليست نفس القدرة المرتفعة التي يمكن أن تتمتع بها اسئلة المقال، كما أن هذه النوعية من الأسئلة لها القدرة على التمييز بين المستويات التحصيلية المختلفة، كما أن استخدام انواع معينة من الأسئلة الاختبارية مثل الأسئلة المقالية وأسئلة الاختيار من متعدد وسائل جيدة لقياس عمق المعرفة، وهذا ما أثبتته دراسة Norbert & Janice (2011) في حين تعارضت نتائج الدراسة الحالية بشان نوعي الأسئلة (اختيار من متعدد، صواب وخطأ) مع نتائج دراسة Tasdemir & Mehmet (2010) التي أظهرت أن أسئلة الاختيار من متعدد ليس لها الأفضلية على اسئلة الصواب والخطأ والمزاوجة في قياس المستوى التحصيلي للطلاب،.. هذا ومن خلال مناقشة النتائج التي تتعلق بأنواع أسئلة الاختبار وعلاقتها بكفاءة الاختبار فقد تبين أن أسئلة المقال يليها أسئلة المزاوجة ثم اسئلة الاختيار

من متعدد ثم أسئلة الصواب والخطأ هي الأسئلة التي يمكن الاعتماد عليها في تحديد الكفاءة النسبية للاختبار التحصيلي البنكي وفقا لأوزان نسبية تتسق مع التأثير والاسهام النسبي لهذه الأنواع من الأسئلة في قدرة التمميز بين مرتفعي ومنخفضي المعدل التراكمي كما بدا من خلال نموذج التحليل

*** نتائج التحليل التمييزي لمتغير تمثيل نطاق القياس:**

أسفرت التحليلات التي اعتمدت على المستويات الثلاث من متغير تميل نطاق القياس (منخفض- متوسط- مرتفع) (كتغيرات ثلاثة مستقلة تمييزية) ومتغير مجموعتي مرتفعي ومنخفضي المعدل التراكمي (كتغير تابع تصنيفي) عن النتائج الموضحة بالجداول التالية:

**جدول (٤٨)**

**جدول الإحصاءات الوصفية لمتغيرات مستوى تمثيل نطاق المتغير المقاس كما يبدو من نتائج التحليل التمييزي**

| المجموعة | متغير تمثيل النطاق | المتوسط الحسابي | الإنحراف المعياري |
|---|---|---|---|
| منخفضي المعدل التراكمي | تمثيل بنسبة ١٠٠% | ٤٧.٤٣٩٧ | ١٢.٤٩٣١٩ |
| | تمثيل بنسبة ٦٦.٦% | ٤٤.٠٠٧١ | ١١.٥٦٠٩٠ |
| | تمثيل بنسبة ٣٣.٣% | ٤٠.٩٠٠٧ | ١١.٠٧٦٩٧ |
| مرتفعي المعدل التراكمي | تمثيل بنسبة ١٠٠% | ٧٤.٤٥٢١ | ٧.٥٩٥٩٠ |
| | تمثيل بنسبة ٦٦.٦% | ٧٣.٦٠٩٦ | ٦.٦٩٥٣٧ |
| | تمثيل بنسبة ٣٣.٣% | ٦٨.٥٣٤٢ | ٦.٣٢٢١٦ |

يتبين من النتائج الموضحة بالجدول (٤٨) وجود فروق ظاهرية بين متوسطات درجات عينتي منخفضي ومرتفعي المعدل التراكمي على أسئلة الاختبار التحصيلي التي تختلف باختلاف نسب تمثيل نطاق

القياس (١٠٠%، ٦٦.٦%، ٣٣.٣%)، حيث بلغ أعلى متوسط القيمة (٧٤.٤٥) تقريبا، وكان لدرجات عينة مرتفعي المعدل التراكمي على كل النطاق، بينما بلغ أدنى متوسط القيمة (٦٨.٦١) تقريبا وكان لعينة منخفضي المعدل التراكمي على (٣٣.٣%) من نطاق القياس..

**جدول (٤٩)**

**جدول اختبار Box's M لدلالة تجانس تباين أو تغاير مجموعتي المرتفعين والمنخفضين بتحليلات نوع المفردة**

| | |
|---|---|
| اختبار Box's M | ٣.٧٨١ |
| مستوى الدلالة | ٠.٠٨٨ |

ويؤكد الجدول (٤٩)، أن قيمة (ف) لاختبار (Box's M) بلغت (٣.٧٨١) وهي قيمة غير دالة، وبالتالي فإن البيانات تستوفي شرط الاعتدالية وتجانس التباين والتغاير.

**جدول (٥٠)**

**معلومات حول خطوات دخول المتغيرات ودلالة كل منها في التحليل التمييزي لمتغيرات نوع المفردات**

| | الخطوة | Tolerance | Sig. of F to Remove | Wilks' Lambda |
|---|---|---|---|---|
| ٢ | تمثيل بنسبة ١٠٠% | ٠.١٢٩ | ٠.٠٠٠ | ٠.٥٧٣ |
| | تمثيل بنسبة ٦٦.٦% | ٠.١٢٩ | ٠.٠٠٠ | ٠.٤٢٧ |
| ٣ | تمثيل بنسبة ١٠٠% | ٠.٠٧٩ | ٠.٠٠٠ | ٠.٣٨٨ |
| | تمثيل بنسبة ٦٦.٦% | ٠.٠٢٦ | ٠.٠٠٠ | ٠.٤١٢ |
| | تمثيل بنسبة ٣٣.٣% | ٠.٠٤٢ | ٠.٠٠٠ | ٠.٣٦٨ |

يتضح من الجدول (٥٠) ترتيب دخول المتغيرات الثلاثة في النموذج، وقد تم ترتيب هذه المتغيرات وفقاً لتأثير كل منها أو نسبة التباين بين المجموعتين الذي يفسره كل منها؛ حيث يتم البدء بالمتغير

الأكثر تأثيراً وهو التمثيل الكامل للنطاق (بنسبة ١٠٠%) ثم تمثيل ثلثي النطاق (بنسبة ٦٦.٦%) ثم تمثيل ثلث النطاق بنسبة (٣٣.٣%).

**جدول (٥١)**

**قيمة الجذر الكامن ومعامل الارتباط القانوني للتحليلات الخاصة بمستويات متغير التخمين**

| الدالة | الجذر الكامن | نسبة التباين % | النسبة التراكمية للتباين % | معامل الإرتباط القانوني |
|---|---|---|---|---|
| ١ | ١.٨٦٢ | ١٠٠.٠ | ١٠٠.٠ | ٠.٨٠٧ |

يتضح من الجدول (٥١) أن قيمة الجذر الكامن بلغت (١.٨٦٢) كما أن قيمة الارتباط القانوني قد بلغت (٠.٨٠٧)، وهو يشير إلى مدى الارتباط بين الدرجات التمييزية وفئات المتغير النوعي أو المجموعات. وبتربيع قيمة هذا المعامل تم الحصول على قيمة مربع إيتا التي بلغت (٠.٦٥)، تقريبا وهو أحد مؤشرات حجم الأثر التي تعتمد على تحديد نسبة التباين المفسر.

**جدول (٥٢)**

**قيمة معامل ويلكس لامدا ودلالته الإحصائية**

| اختبار الدوال | اختبار قيمة Wilks' Lambda | مربع كاي | درجات الحرية | مستوى الدلالة |
|---|---|---|---|---|
| ١ | ٠.٢٤٩ | ٤٤٦.٤٢٨ | ٣ | ٠.٠٠٠ |

ويبين جدول (٥٢) أن قيمة معامل ويلكس لامبدا بلغت (٠.٢٤٩)، وهي دالة عند مستوى (٠.٠١). مما يفيد بصغر كمية التباين غير المفسر في الدرجات التمييزية، مما يفيد بأن أغلب التباين المفسر يعود للمتغيرات التمييزية.

**جدول (٥٣)**

**المعامـلات المعيارية للدالة التمييزيـة**

| المتغيرات التمييزية | المعاملات المعيارية |
|---|---|
| تمثيل بنسبة ١٠٠% | ١.٧٤٢ |
| تمثيل بنسبة ٦٦.٦% | ١.٣٩٤ |
| تمثيل بنسبة ٣٣.٣% | ٠.٣٧٤- |

يتضح من الجدول (٥٣) أن المفردات التي تمثل كل نطاق القياس بنسبة (١٠٠%) بلغ معاملها المعياري القيمة (١.٧٤٢) وهو أكبر من المعامل المعياري للمفردات التي تمثل نسبة (٦٦.٦%) من نطاق القياس والذي بلغت قيمته (١.٣٩٤)، وهو اكبر من المعامل المعياري للمفردات التي تمثل نسبة (٣٣.٣%) من نطاق القياس والذي بلغت قيمته (-١.٣٥٢) وبالتالي فإن تاثير الأسئلة التي تمثل نسبة (١٠٠%) من نطاق القياس مؤثرة في النموذج بدرجة تفوق تأثير الأسئلة الأخرى والتي تمثل نطاق القياس بنسب أقل.

**جدول (٥٤)**

**معامـلات الدالة التمييزيـة**

| المتغيرات التمييزية | المعاملات |
|---|---|
| تمثيل بنسبة ١٠٠% | ٠.٢٥٩ |
| تمثيل بنسبة ٦٦.٦% | ٠.١٢٦ |
| تمثيل بنسبة ٣٣.٣% | ٠.٤٤٨- |
| الثابت | ٤.٥٩٣- |

يوضح الجدول (٥٤) معاملات الدالة التمييزية لكل متغير من المتغيرات التمييزية وهذه القيم هي قيم (أ) للأسئلة التي تمثل (١٠٠%)

من نطاق القياس وبلغت قيمتها (٠.٢٥٨)، (ب) للأسئلة التي تمثل (٦٦.٦%) من نطاق القياس وبلغت قيمتها (٠.١٢٦)، و(ج) للأسئلة التي تمثل (٣٣.٣%) من نطاق القياس وبلغت قيمته (-٠.٤٤٨) بالإضافة على ثابت المعادلة (ث) والذي بلغت قيمته (-٤.٥٩٣) وهي معاملات معادلة الانحدار التي يمكن التنبؤ من خلالها بالمجموعة التي تنتمي إليها الحالات الجديدة.

**جدول (٥٥)**

**إقتران تصنيف عضوية أفراد فئتي منخفضي ومرتفعي التراكمي بالتصنيف المتنبا به من نموذج التحليل**

| المجموعة الأصلية المصنفة في ضوء المعدل التراكمي | المجموعة التنبأ بها من التحليل | | المجموع |
|---|---|---|---|
| | منخفضين | منخفضين | |
| منخفضين | ٢٥٦ | ٢٦ | ٢٨٢ |
| مرتفعين | ٢٠ | ٢٢٤ | ٢٤٦ |

يتضح من جدول (٥٥) أن إثنى وعشرون حالة فقط (١٨ + ٢٠ = ٣٨) كان التنبؤ بها خاطئاً حيث تم تصنيفهم من قبل التحليل في مجموعات خاطئة فقد صنف نموذج التحليل (١٨) حالة من منخفضي المعدل التراكمي بفئة مرتفعي المعدل التراكمي المتوقعة من التحليل، بينما صنفت (٢٠) حالة من مرتفعي المعدل التراكمي بفئة مرتفعي منخفضي الراكمي المتوقعة من قبل التحليل، بينما بقية الحالات (٢٥٦ + ٢٢٤ = ٤٨٠) كان التنبؤ بها صحيحاً، وهذا يشير إلى أن نسبة صحة التنبؤ تبلغ (٤٨٠ ÷ ٥٢٨) = ٩١% تقريباً.

**جدول (٥٦)**

**معامل قوة الاتفاق لكابا بين التصنيف المتنبأ والتصنيف الأصلي لأفراد فئتي منخفضي ومرتفعي المعدل التراكمي**

| اختبار Kappa | القيمة | مستوى الدلالة |
|---|---|---|
| | ٠.٩٠٤ | ٠.٠٠٠ |

وتؤكد قيمة معامل كابا من جدول (٥٦) على فاعلية النموذج في التنبؤ، فقد بلغت (٠.٩٠) تقريبا مما يدل على وجود ارتباط قوي بين اتساق التصنيف المتوقع من النموذج الإحصائي إعتماد على المتغيرات التمييزية المتضمنة به لأفراد عينة البحث مع تصنيفهم الأصلي بمجموعتي منخفضي ومرتفعي المعدل التراكمي.

ويتضح مما سبق من نتائج أن الأسئلة الاختبارية التي تمثل نسبة (١٠٠%) من نطاق القياس تفوقت على الأسئلة الأخرى التي تمثل نسب أقل من نطاق القياس وهي (٦٦.٦%)، (٣٣.٣%) من نطاق القياس في القدرة على التمييز بين مرتفعي ومنخفضي المعدل التراكمي، وقد أكدت هذه النتائج بعض الدراسات التي تناولت العلاقة بين تمثيل المقرر للنطاق السلوكي المستهدف من الاختبار ودقة نتائج القياس.

ومن بين هذه الدراسات دراسة Swiatek & Mary (2005) التي أظهرت أن الطلاب المتفوقون تفوق أدائهم على اختبارات الدراسة الممثلة لنطاق القياس المستهدف بشكل جيد عن أداء الطلاب العاديون، بينما لم يختلف جوهريا عن اداء العاديون في الاختبارات التي لا يمثل النطاق السلوكي المستهدف بشكل جيد. كما أكدت أيضاً دراسة Hudson, Lewis & Stichter (2011) نفس النتائج حيث توصلت تلك الدراسة

إلى بعض المؤشرات التي يمكن تميز جودة الاختبار من بينها تغطية أسئلة الاختبار لجميع موضوعات وفقرات المقرر الدراسي.

وقد أشارت الدراسة أن أسئلة الاختبار ينبغي تعكس الأهداف التعليمية المرتبطة بنسبة يجب أن تتجاوز النسبة (٦٠%) من جملة الأهداف التعليمة، كما اتسقت هذه النتائج مع النتائج التي أشارت إليها دراسة Tyrone & David (2011) حيث أظهرت أن الاختبارات المقننة التي تعدها الهيئات المتخصصة هي الاختبارات التي يمكن أن تقيس التعلم بصورة أدق واكثر موضوعية نظراً لأنها اختبارات مقننة تحتوي مفردات تمثل السلوك المتعلم بصورة ادق وأشملز

أما الاختبارات التحصيلية التي يتم انتاجها محليا هي اخبتارات لاتتمتع بخاصية تمثيل المجال المستهدف من القياس، وتبين من مناقشة النتائج السابقة أن المفردات التي تمثل النطاق السلوكي المستهدف من القياس بصورة أشمل ترفع من كفاءة الاختبار بما يؤكد إمكانية الاعتماد على معيار شمول عينة اسئلة الاختبار للنطاق السلوكي المستهدف قياسه في تقييم كفاءة الاختبار التحصيلي البنكي.

*** نتائج التحليل التمييزي للمستويات المعرفية:**

أسفرت التحليلات التي اعتمدت على المستويات الثلاث من متغير المستويات المعرفية (تذكر- فهم- تطبيق- عليا) (كتغيرات مستقلة تمييزية) ومتغير مجموعتي مرتفعي ومنخفضي المعدل التراكمي (كتغير تابع تصنيفي) عن النتائج الموضحة بالجداول التالية:

## جدول (٥٧)

## جدول الإحصاءات الوصفية لمتغيرات المستوى المعرفي

## كما يبدو من نتائج التحليل التمييزي

| المجموعة | متغير تمثيل النطاق | المتوسط الحسابي | الإنحراف المعياري |
|---|---|---|---|
| منخفضي المعدل التراكمي | تذكر | ١٨.٥٦٧٤ | ٥.٦٣٠٧٦ |
| | فهم | ١٧.٠٠٣٥ | ٤.٨٣٣٥٣ |
| | تطبيق | ٦.٦٠٦٤ | ٢.٥٧٢٣١ |
| | عليا | ٤.٤٢٩١ | ١.٧٥٦٦٢ |
| مرتفعي المعدل التراكم | تذكر | ٢٧.١٥٧٥ | ٢.٥٥٦١٠ |
| | فهم | ٢٥.١٧١٢ | ٣.٤٤٠٨٥ |
| | تطبيق | ١٢.٣٣٥٦ | ١.١٧٠١٩ |
| | عليا | ٨.٣٨٣٦ | ١.٦٧٤٣٩ |

يتبين من النتائج الموضحة بالجدول السابق (٥٧) وجود فروق ظاهرية بين متوسطات درجات عينتي منخفضي ومرتفعي المعدل التراكمي على أنواع الأسئلة التي تختلف باختلاف المستوى المعرفي (تذكر، فهم، تطبيق، مستويات عليا).

فقد بلغ أعلى متوسط القيمة (٢٧.١٥٧٥) وكان لدرجات عينة مرتفعي المعدل التراكمي على الأسئلة التي تندرج تحت مستوى التذكر. بينما بلغ أدنى متوسط القيمة (٤.٤٢٩١) وكان لعينة منخفضي المعدل التراكمي على الأسئلة التي تندرج تحت تحت المستويات المعرفية العليا.

**جدول (٥٨)**

**جدول اختبار Box's M لدلالة تجانس تباين أو تغاير مجموعتي المرتفعين والمنخفضين**

| | |
|---|---|
| اختبار Box's M | ٧.١١٢ |
| مستوى الدلالة | ٠.٢١٤ |

ويؤكد الجدول السابق (٥٨)، أن قيمة (ف) لاختبار (Box's M) بلغت (٧.١١٢) وهي قيمة غير دالة، وبالتالي فإن البيانات تستوفي شرط الاعتدالية وتجانس التباين والتغاير.

**جدول (٥٩)**

**معلومات حول خطوات دخول المتغيرات ودلالة كل منها في التحليل التمييزي**

| الخطوة | | Tolerance | Sig.F to Remove | Wilks' Lambda |
|---|---|---|---|---|
| ٢ | التطبيق مستوى | ٠.٨٠٨ | ٠.٠٠٠ | ٠.٤٥٨ |
| | عليا مستويات | ٠.٨٠٨ | ٠.٠٠٠ | ٠.٣٩٥ |

الجدول (٥٩) يوضح الخطوة الأخيرة لترتيب دخول المتغيرين في النموذج. وقد تم ترتيب هذه المتغيرات وفقاً لتأثير كل منها أو نسبة التباين بين المجموعتين الذي يفسره كل منها؛ حيث يتم البدء بالمتغير الأكثر تأثيراً وهو مستوى التطبيق ثم المستويات العليا.

**جدول (٦٠)**

**قيمة الجذر الكامن ومعامل الارتباط القانوني**

| الدالة | الجذر الكامن | نسبة التباين % | النسبة التراكمية للتباين % | معامل الإرتباط القانوني | الدالة |
|---|---|---|---|---|---|
| ١ | a١.٩٠١ | ١٠٠ | ١٠٠ | ٠.٨١ | ٠.٦٦ |

يتضح من الجدول (٦٠) أن قيمة الجذر الكامن بلغت (١.٩٠١) كما أن قيمة الارتباط القانوني قد بلغت (٠.٨١)، وهو يشير إلى قوة الارتباط بين الدرجات التمييزية وفئات المتغير النوعي أو المجموعات، كما أن قيمة مربع إيتا التي بلغت (٠.٦٦)، وهو نسبة التباين المفسر بين مجموعتي المرتفعين والمنخفضين والتي تعزى للمتغيرات التمييزية.

**جدول (٦١)**

**قيمة معامل ويلكس لامدا ودلالته الإحصائية**

| اختبار الدوال | اختبار قيمة Wilks' Lambda | مربع كاي | درجات الحرية | مستوى الدلالة |
|---|---|---|---|---|
| ١ | ٠.٢٤٥ | ٤٥٢.٦٦٥ | ٢ | ٠.٠٠٠ |

ويبين الجدول (٦١) أن قيمة معامل ويلكس لامبدا بلغت (٠.٢٤٥)، وهي دالة عند مستوى (٠.٠١). مما يفيد بصغر كمية التباين غير المفسر في الدرجات التمييزية والتي لاتعزى للمتغيرات التمييزية.

**جدول (٦٢)**

**المعامـلات المعيارية للدالة التمييزيـة**

| المتغيرات التمييزية | المعاملات المعيارية |
|---|---|
| مستوى التطبيق | ٠.٦٨٤ |
| مستويات عليا | ٠.٤٨٩ |

ويتضح من الجدول (٦٢) الذي يوضح معاملات المعيارية للدالة التمييزية، أن القيم الواردة للمعاملات تبين أن الاختبار مستوى التطبيق بلغ معامله المعياري القيمة (٠.٦٨٤) وهو اكبر من المعامل المعياري للمستويات المعرفية العليا والذي بلغت قيمته (٠.٤٨٩) وبالتالي يتضح أن مستوى التطبيق كان مؤثرا في النموذج بدرجة تفوق تأثير المستويات

المعرفية العليا، وهذه المعاملات مفيدة في المقارنة بين أثر كل متغير كمي في النموذج.

**جدول (٦٣)**

**معامــلات الدالة التمييزيــة**

| المتغيرات التمييزية | المعاملات |
|---|---|
| مستوى التطبيق | ٠.٣١١ |
| مستويات عليا | ٠.٢٨٣ |
| الثابت | -٤.٢٩٩ |

يتبين من الجدول (٦٣) معاملات الدالة التمييزية لكل متغير من المتغيرات التمييزية (المستقلة) وهذه القيم هي قيم (أ) لمتغير مستوى التطبيق وبلغت قيمتها (٠.٣١١)، و(ب) لمتغير المستويات المعرفية العليا وبلغت قيمته (٠.٢٨٣)، بالإضافة على ثابت المعادلة (ج) والذي بلغت قيمته (-٤.٢٩٩) وهي القيم اللازمة لمعادلة الانحدار التي تمكننا من التنبؤ بالمجموعة التي تنتمي إليها الحالات الجديدة، كما سيرد لا حقاً.

**جدول (٦٤)**

**إقتران تصنيف عضوية أفراد فئتي منخفضي ومرتفعي التراكمي بالتصنيف المتنبا به من نموذج التحليل**

| المجموعة الأصلية المصنفة في ضوء المعدل التراكمي | المجموعة التنبأ بها من التحليل | | المجموع |
|---|---|---|---|
| | منخفضين | مرتفعين | |
| منخفضين | ٢٦٣ | ١٩ | ٢٨٢ |
| مرتفعين | ٥ | ٢٤١ | ٢٤٦ |

يتضح من جدول (٦٤) أن أربع وعشرون حالة (٥ + ١٩= ٢٤) كان التنبؤ بها خاطئاً حيث تم تصنيفهم من قبل التحليل في مجموعات

خاطئة فقد صنفت (٥) حالات من مرتفعي المعدل التراكمي بفئة منخفضي المعدل التراكمي، بينما صنف عدد (١٩) حالة منخفضي المعدل التراكمي بفئة مرتفعي المعدل الراكمي، بينما بقية الحالات (٢٦٣ + ٢٤١ = ٥٠٤) كان التنبؤ بها صحيحاً، وهذا يشير إلى أن نسبة صحة التنبؤ تبلغ (٥٠٤ ÷ ٥٢٨ )= ٠.٩٥ تقريباً.

**جدول (٦٥)**

**معامل قوة الاتفاق لكابا بين التصنيف المتنبأ به والتصنيف الأصلي لأفراد فئتي منخفضي ومرتفعي المعدل التراكمي**

| اختبار Kappa | القيمة | مستوى الدلالة |
|---|---|---|
| | ٠.٩٠٣ | ٠.٠٠٠ |

وتؤكد قيمة معامل كابا من جدول (٦٥) على فاعلية النموذج في التنبؤ، فقد بلغت (٠.٩٠) مما يدل على وجود ارتباط قوي جداً بين النتيجة الفعلية والنتيجة المتنبأ بها. مما يشير إلى قوة اتساق التصنيف المتوقع من النموذج الإحصائي إعتماد على المتغيرات التمييزية المتضمنة به لأفراد عينة البحث مع تصنيفهم الأصلي بمجموعتي منخفضي ومرتفعي المعدل التراكمي.

يتضح مما سبق من نتائج، أن كل من أسئلة مستوى التطبيق وأسئلة المستويات العليا وفقا لتصنيف بلوم كانتا أكثر تأثيرا في نموذج التحليل، وبالتالي كانتا أكثر قدرة على التمييز بين عينتي مرتفعي ومنخفضي المعدل التراكمي، كما بدا من خلال تلك النتائج أن تأثير أسئلة الاختبار التي تقيس مستوى التطبيق في التمييز بين مجموعتي مرتفعي ومنخفضي المعدل التراكمي أكثر من أسئلة الاختبار من المستويات العليا (المعرفية)،

مما يفيد بأن هذين النوعين من المستويات المعرفية بتصنيف بلوم يمكن الاعتماد عليهما في تقييم كفاءة الاختبار التحصيلي، وهذه النتائج أكدتها دراسة Aaron, Irina & David (2005) التي أظهرت أن البنود الاختبارية التي تقيس المستويات العليا بتصنيف بلوم وهي التي تتعلق بمستويات التحليل والتركيب والتقويم وهي التي ترتبط بالتقويم الحقيقي لعملية التعلم.

كما أكدت هذه النتيجة أيضاً دراسة Tavakol, et.al.( 2011) التي أظهرت أن توظيف الطالب للمعلومات في معالجة بعض المواقف الجديدة مؤشرا يمكن أن تساهم في إعداد اختبار يمكن أن يستخدم بكفاءة في تقويم الأداء الأكاديمي، ولهذين المستوين بتصنيف بلوم أهمية في تحديد كفاءة الاختبار التحصيلي فقد أشارت دراسة Nilay (2007) أن المعرفة الإجرائية ترتبط بالأسئلة التي تندرج تحت مستوى التطبيق، وأن مهارات التفكير الناقد ترتبط بالأسئلة التي تندرج تحت مستويي التركيب والتقويم مما يؤكد أن هذين المستويين (مستوى التطبيق، المستويات العليا) بتصنيف بلوم ينبغي أن تشكل الأساس الذي ينبغي أن ترتكز إليه عملية تقييم الكفاءة النسبية للاختبار، وقد أكدت تلك النتيجة أيضاً دراسة Kucuk & Walters (2009) التي أوضحت أن المستويات المعرفية العليا وفقا لتصنيف بلوم هي مؤشرا جيدا على كفاءة الاختبار، كما أكدت تلك النتيجة دراسة James (2011) التي أثبتت فعالية نموذج مقترح للتقييم يعتمد على تصنيف بلوم للأهداف في المجال المعرفي بصفة عامة وبصفة خاصة مستوى التطبيق والمستويات العليا من تصنيف بلوم كما تتاكد كذلك تلك النتيجة مع ما أشارت إله دراسة Bowers & Alex (2011) من أن اغلب الاختبارات

التحصيلية بالجامعات الأمريكية كانت لا تتعلق بقدرة المتعلم على توظيف ماتم تعلمه بمواقف جديدة لم يسبق له أن مر بها من قبل كما أنها لا تتضمن قياسات خاصة بالمستويات العليا من تصنيف بلوم للأهداف.

هذا ويمكن تفسير تفوق الأسئلة التي تندرج تحت مستوى التطبيق في قدرتها على التمييز بين مجموعتي مرتفعي ومنخفضي المعدل التراكمي على الأسئلة التي تندرج تحت المستويات العليا المعرفية باهتمام المؤسسات التعليمية الجامعية بالتقويم المعتمد على الأداء وهو مايسمى بالتقويم الحقيقي Authentic assessment وهذا ما اكدته دراسة Newhouse (2011) والتي أظهرت أن تقويم التعلم الحقيقي للطلاب يتعلق بالتطبيقات الخاصة بموضوعات التعلم، وهذا بالفعل ما تأكد من خلال نتائج الدراسة الحالية والتي اظهرت أهمية قياس الأهداف التعليمية التي تندرج مستوى التطبيق وكذلك المستويات المعرفية العليا. بما يؤكد أن كفاءة الاختبار التحصيلي يمكن أن تقاس من خلال مدى شموله على أسئلة تترد تحت تلك المستويات.

## ثالثاً: عرض ومناقشة نتائج الإجابة عن السؤال الثالث من أسئلة الدراسة:

لقد سعت الدراسة إلى الإجابة عن السؤال الثالث والذي ينص على: هل يختلف رتب أفراد عينة الدراسة وفقا لمعدلهم التراكمي مع كل من رتب تقديرات قدراتهم المشتقة من استجاباتهم على أسئلة الاختبار ورتب درجاتهم الحقيقية على نفس أسئلة الاختبار التي صنفت في ضوء متغيرات الدراسة (الصعوبة- التمييز- دالة

المعلومات– التخمين– نوع المفردات– نسبة تمثيل نطاق المتغير المقاس– المستوى المعرفي) بمستوياتها المختلفة.؟ اعتماداً على إجراءين مختلفين أحدهما يتعلق بالشق الأول من السؤال والآخر يتعلق بالشق الثاني كالآتي:

**إجراء الإجابة عن الشق الأول:**

- تحليل استجابات العينة الكلية على أسئلة اختبار الدراسة والتي صنفت في ضوء المتغيرات قيد البحث بمستوياتها المختلفة اعتمادا على استخدام نموذج التقدير الجزئي للحصول على تقديرات لقدرات الأفراد مقدرة بوحدة اللوجيت وقد حولت هذه التقديرات بوحدة أخرى هي الوات للتخلص من التقديرات السالبة والعشرية، ثم أوجدت الدراسة معامل ارتباط الرتب لسبيرمان بين تقديرات القدرة للعينة المقدرة بوحدة الوات والمعدل التراكمي لهم.

**إجراء الإجابة عن الشق الثاني:**

- تقدير الدرجات الحقيقية، بصيغة نانلي Nunnally, & (Bernstein,1994) لكل فرد من أفراد عينة الدراسة والتي اشتقت من درجاتهم الخام الحاصلون عليها من استجاباتهم لمفردات اختبار الدراسة المصنفة تبعا للمتغيرات قيد البحث بمستوياتها المختلفة، ثم أوجدت الدراسة معامل ارتباط الرتب لسبيرمان بين الدرجات الحقيقية للعينة والمعدل التراكمي لهم.

هذا وقد أسفر الإجراءان السابقان عن البيانات الموضحة بالجدول التالي:

## جدول (٦٦)

## معاملات إرتباط رتب المعدل التراكمي للعينة بكل من رتب تقديرات القدرة والدرجات الحقيقية والتي اشتقت من استجاباتهم على أسئلة اختبار الدراسة المصنفة في ضوء متغيرات الدراسة بمستوياتها المختلفة

| المتغير | مستوى المتغير | معامل سبيرمان لإرتباط الرتب بين | |
|---|---|---|---|
| | | تقديرات القدرة والمعدل التراكمي ر(ق، م) | الدرجة الحقيقية والمعدل التراكمي (ح، م) |
| الصعوبة | منخفض | ٠.٧٩٢** | ٠.٧٥٤** |
| | متوسط | ٠.٩٤٣** | ٠.٨٥٠** |
| | مرتفع | ٠.٨١٧** | ٠.٧٩١** |
| قوة التمييز | منخفض | ٠.٦٢٥** | ٠.٧٥٥** |
| | متوسط | ٠.٩٠١** | ٠.٨٠٧** |
| | مرتفع | ٠.٧٣١** | ٠.٧٧٧** |
| أقصى معلومات | منخفض | ٠.٦١٧** | ٠.٧٣٣** |
| | متوسط | ٠.٨١٠** | ٠.٨٢٠** |
| | مرتفع | ٠.٩٦٥** | ٠.٨١٤** |
| التخمين | منخفض | ٠.٩٢٢** | ٠.٧٨٨** |
| | متوسط | ٠.٧٥٣** | ٠.٧٤١** |
| | مرتفع | ٠.٥٥٥** | ٠.٧٠٥** |
| نوع المفردة | اختيار من متعدد | ٠.٧٣٣** | ٠.٧٠٠** |
| | صواب وخطأ | ٠.٦٣٣** | ٠.٥٩٦** |
| | مزاوجة | ٠.٨١٦** | ٠.٧٦٨** |
| | إكمال | ٠.٥٦١** | ٠.٥٨٦** |
| | إجابة قصيرة | ٠.٤٠٧** | ٠.٣٤٧** |
| | مقال | ٠.٨٥٦** | ٠.٧٢٢** |
| تمثيل نطاق متغير القياس | كامل النطاق | ٠.٩٢٠ | ٠.٩١١ |
| | ثلثا النطاق | ٠.٦٦٥ | ٠.٦٨٨ |
| | ثلث النطاق | ٠.٤٤٣ | ٠.٤٩١ |
| المستوى المعرفي | تذكر (معرفة) | ٠.٦٦٢** | ٠.٦٤٤** |
| | فهم | ٠.٧٩٢** | ٠.٧٠٥** |
| | تطبيق | ٠.٩٥٤** | ٠.٨٤٢** |
| | عليا | ٠.٨٤٩** | ٠.٧٨٩** |

يتضح من جدول(٦٦) فيما يتعلق بمتغير الصعوبة وجود اتساق بين معاملات إرتباط الرتب المتناظرة، وهي المعاملات ر(ق، م) المشتقة من إرتباط تقديرات القدرة (ق) والمعدل التراكمي (م)، والأخرى ر(ح، م)المشتقة من إرتباط الدرجات الحقيقية (ح) والمعدل التراكمي(م)حيث تبين أن أعلى معاملات إرتباط لكل من ر(ق، م)، ر(ح، م) كان للمستوى المتوسط من الصعوبة يلية المستوى المرتفع ثم المنخفض، وتبدو هذه النتيجة أمرا منطقياً لأن المستوى المتوسط من صعوبة الأسئلة يؤدي إلى مزيد من كفاءة الاختبار وهذا ماتؤكده العديد Aaron,et.al. (2005)، Jandaghi & Walters (2009), Shateria (2008) من الدراسات والأدبيات، كما أن المستوى المتوسط من صعوبة البنود يرتبط بالدرجة الحقيقية Tavakol, et.al. (2011)، كما يتضح من جدول (٦٦) فيما يتعلق بمتغير التمييز وجود اتساق معاملات إرتباط الرتب المتناظرة، وهي ر(ق، م)، ر(ح، م) للمستوى المتوسط من التمييز يلية المستوى المرتفع ثم المنخفض، وهذا يبدوا أمراً منطقياً لأن المستوى المتوسط من قوة تمييز الأسئلة يؤدي إلى مزيد من كفاءة الاختبار وقد أكدت هذه النتيجة بعض الدراسات مثل دراسة Clifton & Schriner (2010)، Tasdemir, Mehmet (2010)، Ryan, Peter & Elizabeth (2010)، كما أن المستوى المتوسط من متغير التمييز يمكن الاستناد إليه كأحد المعايير التي يمكن استخدامها في تقييم الكفاءة النسبية للاختبار. وفي تقدير مستوى الأداء الحقيقي للطالب وهذا ما أشارت إليه أيضاً كل من دراستي Tavakol, et.al. (2011)، Richard, (2001). ويتضح أيضاً من جدول (٦٦) فيما يتعلق بمتغير أقصى معلومات وجود حالة من عدم

الاتساق فيما يتعلق بمعاملات الإرتباط المتناظرة، وهي ر(ق، م)، ر(ح، م) حيث جاءت أعلى قيمة لمعامل ر(ق، م) للمستوى المرتفع من المعلومات يلية المستوى المتوسط ثم المنخفض، بينما جاءت أعلى قيمة لمعامل ر(ح، م) للمستوى المتوسط من المعلومات يليها المستوى المرتفع ثم المنخفض ولكن مايلاحظ مع هذا المعامل هو أن الفارق بين معاملي المستويين المتوسط والمرتفع بلغ القيمة (٠.٠٠٦) وهو يبدو فارقا ضئيلاً للغاية، وبالتالي يمكن اعتبار أن مستوى المعلومات المتوسط تقريبا مشابه إلى حد كبير للمستوى المعلومات المرتفع من حيث النتائج، وهذه النتائج أكدتها الأطر النظرية والأدبيات التي أفادت بأن المفردة التي تمتلك أعلى قيمة للمعلومات أفضل من مفردات أخرى تمتلك معلومات أقل عند نفس المستوى من القدرة مما يعطي تقديرات ادق للقدرة المستهدف قياسها من تلك المفردات وهذا ما اكدته العديد من الدراسات(Siang & Fritz ،Kingsbury & Zara (2002) (2006)، Edward (2010)، (Marinagi, & Kaburlasos, (2010)، Altaf, et.al (2011)، Lihua Yao (2011)، كما يتضح من جدول(٦٦) أيضاً فيما يتعلق يمتغير التخميين اتساق معاملات الإرتباط المتناظرة، وهي ر(ق، م)، ر(ح، م) حيث جاءت أعلى قيمة للمعاملات للمستوى المنخفض من التخمين يلية المستوى المتوسط ثم المرتفع، وهذا يبدوا أمرا منطقياً لأن المستوى المنخفض من التخمين الأسئلة يؤدي إلى مزيد من كفاءة الاختبار، مما يعني أن المستويات المتزايدة من هذا المتغير أثرت في تصنيف أفراد عينة البحث بما يتفق مع تصنيفهم في ضوء معدلهم التراكمي، وقد تأكدت هذه النتيجة من خلال بعض الأدبيات التي أشارت إلى أن سلوك التخمين يؤثر على

تقديرات القدرة للمختبرين Richard (2001)، Chang, et (2011)، Schurmeier, et.al(2011). ويتضح من جدول(٦٦) فيما يتعلق بمتغير نوع المفردات وجود اتساق بين معاملات الإرتباط المتناظرة، وهي المعاملات ر(ق، م) ر(ح، م) حيث تبين أن أعلى معاملات لإرتباط لكل منهما كان لأسئلة المقال تليها أسئلة المزاوجة ثم الاختيار من متعدد ثم الصواب والخطأ، ثم الإكمال ثم الإجابة القصيرة، وتبدو هذه النتائج منطقية نظرا لأهمية أسئلة المقال في قياس المستويات المعرفية العليا وهذا ما أشارت إليه دراسة Clifton & Schriner (2010) ، Schurmeier, et al (2010)، Hicks, et al (2010)، Norbert & Janice (2011) كما أن ارتفاع قيم معاملات ارتباط الرتب لكل من الدرجات الحقيقية لأسئلة الاختيار من متعدد والمزاوجة مع المعدل التراكمي يفيد بضرورة احتواء الاختبار التحصيلي على هذين النوعين من الأسئلة حيث أكدت دراسة Tasdemir & Mehmet, (2010) على أن إحتواء الاختبار على اسئلة متنوعة من الاختيار من متعدد والمزاوجة له آثار إيجابية على ثبات الاختبار، كما أن اختبار الاختيار من متعدد ليس له الأفضلية على المزاوجة في قياس المستوى التحصيلي للطلاب.

وهذا ما أكدته أيضاً دراسة كل من Jandaghi, & Shateria (2008)، Walters, Kucuk (2009) ، حيث أظهرت إلى نوعية المفردات التي يعتمد عليها الاختبار والتي تبتعد إلى حد ما عن التخمين مؤشراً على كفاءة الاختبار.

أما أسئلة الإكمال فقد يفسر معاملها المتدني بأنها قد ترتبط بالمستويات المعرفية الدنيا دون العليا نظرا لأن اغلب هذه الأسئلة ترد من نصوص علمية تنقصها بعض الكلمات (نصار، ٢٠١٠:

٢١٤)، مما يجعلها تتعلق بمستوى التذكر أكثر من أي مستوى آخر.

كما يتضح من جدول (٦٦) فيما يتعلق بمتغير تمثيل نطاق المتغير المقاس وجود اتساق بين معاملات الإرتباط المتناظرة، وهي المعاملات ر(ق، م) والأخرى ر(ح، م) حيث تبين أن أعلى معاملات لإرتباط لكل من ر(ق، م)، ر(ح، م) كان للتمثيل التام لنطاق المتغير المستهدف من القياس (بنسبة ١٠٠%) يلية التمثيل بنسبة (٦٦.٦%) ثم التمثيل بنسبة (٣٣.٣%) وتبدو هذه النتيجة أمرا منطقياً لأن تمثيل الأسئلة لكامل نطاق المتغير المستهدف قياسه يؤدي إلى مزيد من كفاءة الاختبار والتي تتجسد في قدرة الاختبار في التمييز بين مستوياتهم الحقيقية على خلفية المعدل التراكمي لهم وهذا ماتؤكده دراسات Lewis & Stichter (2011)، Switek & May, Hudson (2011)، Tyrone & David (2011).

ويتضح من أيضاً من جدول (٦٦) فيما يتعلق بمتغير المستوى المعرفي وجود اتساق بين معاملات الإرتباط المتناظرة، وهي المعاملات ر(ق، م) والأخرى ر(ح، م) حيث تبين أن أعلى معاملات لإرتباط لكل من ر(ق، م)، ر(ح، م) كان لمستوى التطبيق يلية المستويات العليا يلي ذلك مستوى الفهم ثم مستوى التذكر وهذه النتائج أكدتها العديد من الدراسات(2005) Aaron, Irina & David، Tavakol, et.al. (2011)، Nilay (2007)، Kucuk & James (2011)، Walters (2009).

**رابعاً: عرض ومناقشة نتائج الإجابة عن السؤل الرابع من أسئلة الدراسة:**

لقد اعتمدت إجراءات الدراسة في الإجابة عن السؤال الرابع والذي ينص على: هل يمكن اشتقاق معايير كمية يمكن استخدامها في تقييم الكفاءة النسبية للاختبارات التحصيلية البنكية في ضوء متغيرات الدراسة؟ على تحديد أوزان نسبية لمستويات المتغيرات قيد البحث في ضوء دلالة قوتها التمييزية بين مجموعتي الدراسة من مرتفعي ومنخفضي المعدل التراكمي التي تم الكشف عنها في ضوء نتائج المعاملات المعيارية للمتغيرات التمييزية التي اسفر عنها أسلوب التحليل التمييزي المستخدم بالدراسة، وكذلك في ضوء قوة اتفاق ترتيب أفراد عينة الدراسة في ضوء معدلهم التراكمي مع ترتيب تقديرات القدرة لهم وايضاً ترتيب درجاتهم الحقيقة الحاصلون عليها بموجب استجاباتهم لأسئلة اختبار الدراسة والتي صنفت وفقا لمتغيرات الدراسة بمستوياتها مختلفة، وذلك إعتمادا على معامل ارتباط الرتب لسبيرمان وفيما يلي بيان هذه الإجراءات.

**[١] تحديد درجات موزونة للمتغيرات البحثية ذات الدلالة التمييزية:**

وذلك باتباع الخطوات التالية:

- تحديد المتغيرات التمييزية التي ظهر لها دلالة تمييزية بين مجموعتي مرتفعي ومنخفضي التراكمي وأدخلت بالنموذج الإحصائي.
- تحديد قيمة المعاملات المعيارية للمتغيرات التمييزية (بغض النظر عن الإشارة) والتي تدل على نسبة اسهامها في التمييز بين مجموعتي التنبؤ. (أبو علام، ٢٠٠٩: ٣١٢) وسوف تحدد الدرجات التمييزية كما بالجدول التالي:

**والجدول (٦٧)**

**يوضح الدرجات الموزونة لكل القيم المحتملة للمعاملات المعيارية**

| المعامل المعياري | ٠- ١٠ | ١٠- ١٥ | ١٥- ٢٠ | ٢٠- ٢٥ | ٢٥- ٣٠ | ٣٠- ٣٥ | ٣٥- ٤٠ | ٤٥- ٥٠ | ٥٠- ٥٥ | ٥٥-٦٠ |
|---|---|---|---|---|---|---|---|---|---|---|
| الدرجة لموزونة | ٥ | ١ | ١.٥ | ٢ | ٢.٥ | ٣.٥ | ٤ | ٤.٥ | ٥ | ٥.٥ |
| المعامل المعياري | ٦٠- ٦٥ | ٦٥- ٧٠ | ٧٠- ٧٥ | ٧٥- ٨٠ | ٨٠- ٨٥ | ٩٥- ٩٠ | ٩٠- ٩٥ | ٩٥- ١٠٠ | ١- ١.٥ | ≤ ١.٥ |
| الدرجة لموزونة | ٦ | ٦.٥ | ٧ | ٧.٥ | ٨ | ٨.٥ | ٩ | ٩.٥ | ١٠ | ١.٥ |

**[٢] تحديد درجات موزونة لقوة معاملات الاتفاق (معاملات إرتباط الرتب):**

يمكن تحديد درجات موزونة مقترحة لقوة معاملات الاتفاق (معاملات إرتباط الرتب) كما بالجدول التالي:

**جدول (٦٨)**

**الدرجات الموزونة المقترحة لكل القيم المحتملة لمعاملات ارتباط الرتب**

| معامل الارتباط | الفترة التي ينتمي إليها معامل إرتباط الرتب لسبيرمان | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | ٦٠- ٦٥ | ٦٥- ٧٠ | ٧٠- ٧٥ | ٧٥- ٨٠ | ٨٠- ٨٥ | ٨٥- ٩٠ | ٩٠- ٩٥ | ٩٥- ١٠٠ |
| الدرجة الموزونة | ٠.٥ | ١ | ١.٥ | ٢ | ٢.٥ | ٣ | ٣.٥ | ٤ |

وبالتعويض عن الدرجات الموزونة المقترحة لكل من المعاملات المعيارية ومعاملات إرتباط الرتب يمكن إيجاد مجموع الدرجات الموزونة لكل مستوى محدد من مستويات كل متغير بحثي والجدول التالي يوضح مجموع الدرجات المقترحة الموزونة لأي مستوى من مستويات كل متغير من متغيرات الدراسة:

## جدول (٦٩)

## مجموع الدرجات الموزونة المقترحة لكل مستوى محدد من مستويات المتغير البحثي

| المتغير | مستوى المتغير | المعاملات المعيارية | النقاط الموزونة | | | | مجموع الدرجات الموزونة لمستويات المتغير | نسب توزيع الدرجات الموزونة لكل مستوى إلى مجموع الدرجات الموزونة |
|---|---|---|---|---|---|---|---|---|
| | | | ع | ر(ق، م) | (ح، م) | ع +ر(ق، م)+ر(ح، م) | | |
| الصعوبة | منخفض | -٠.٦١٨ | - | ٢ | ٢ | ٤ | ٣٣ | %٢٨ |
| | متوسط | ١.٠٠٨ | ١٠ | ٣.٥ | ٣ | ١٦.٥ | | %٥٠ |
| | مرتفع | ٠.٨٠٧ | ٨ | ٢.٥ | ٢ | ١٢.٥ | | %٣٨ |
| قوة التمييز | منخفض | ٠.٨١٣ | ٨ | ٠.٥ | ٢ | ١٠.٥ | ٣٦.٥ | %٢٨ |
| | متوسط | ٠.٨٦١ | ٨.٥ | ٣.٥ | ٢.٥ | ١٤.٥ | | %٤٠ |
| | مرتفع | ٠.٨٣٣ | ٨ | ١.٥ | ٢ | ١١.٥ | | %٣٢ |
| أقصى معلومات | منخفض | - | - | ٠.٥ | ١.٥ | ٢ | ٢٤.٥ | %٨ |
| | متوسط | ٠.٧٦٧ | ٧.٥ | ٢.٥ | ٢.٥ | ١٢.٥ | | %٥١ |
| | مرتفع | ٠.٣٠٩ | ٣.٥ | ٤ | ٢.٥ | ١٠ | | %٤١ |
| التخمين | منخفض | ٠.٤٨٥ | ٤.٥ | ٣.٥ | ٣ | ١١ | ٢١.٥ | %٥١ |
| | متوسط | ٠.٣٣١ | ٣.٥ | ٢ | ١.٥ | ٧ | | %٣٣ |
| | مرتفع | ٠.٢٩٧ | ٢.٥ | ٠ | ١ | ٣.٥ | | %١٦ |

| المتغير | مستوى المتغير | المعاملات المعيارية | النقاط الموزونة | | | | مجموع الدرجات الموزونة لمستويات المتغير | نسب توزيع الدرجات الموزونة لكل مستوى إلى مجموع الدرجات الموزونة |
|---|---|---|---|---|---|---|---|---|
| | | | ع | ر(ق، م) | (ح، م) | ع +ر(ق، م)+ر(ح، م) | | |
| نوع المفردة | اختيار من متعدد | ٠.٣٥٢ | ٤ | ١.٥ | ١.٥ | ٧ | ٣٧.٥ | %١٨ |
| | صواب وخطأ | ٠.١٩٢ | ١.٥ | ٠.٥ | ٠ | ٢ | | %٥ |
| | مزاوجة | ٠.٥٣٩ | ٥ | ٢.٥ | ٢ | ٩.٥ | | %٢٥ |
| | إجابة قصيرة | -٠.١٩١ | ١.٥ | ٢ | ٢ | ٥.٥ | | %١٤ |
| | مقال | ٠.٥٥١ | ٥.٥ | ٣ | ١.٥ | ١٠ | | %٢٦ |
| | إكمال | - | ٠ | ٢.٥ | ١.٥ | ٤ | | %١٠ |
| تمثيل نطاق متغير القياس | نسبة تمثيل النطاق (١٠٠%) | ١.٧٢٤ | ١٠.٥ | ٣.٥ | ٣.٥ | ١٧.٥ | ٣٣ | %٥٣ |
| | نسبة تمثيل النطاق (٣٣.٣%) | -٠.٣٧٤ | ٣.٥ | ٠ | ٠ | ٣.٥ | | %١١ |
| | نسبة تمثيل النطاق (٦٦.٦%) | ١.٣٩٤ | ١٠ | ١ | ١ | ١٢ | | %٣٦ |
| المستوى المعرفي | تذكر (معرفة) | - | - | ١ | ٠.٥ | ١.٥ | ٢٢.٥ | %٧ |
| | فهم | - | - | ٢ | ١.٥ | ٣.٥ | | %١٥ |
| | تطبيق | ٠.٦٨٤ | ٦.٥ | ٤ | ٢.٥ | ١٣ | | %٥٨ |
| | عليا | ٠.٤٨٩ | ٤.٥ | ٢.٥ | ٢ | ٤.٥ | | %٢٠ |

ويقترح توزيع مجموع ونسب الدرجات الموزونة المقترحة على مستويات كل متغير من متغيرات البحث كما يلي:

- توزيع الدرجات الموزونة على مستويات متغير الصعوبة لمفردات الاختبار = ٤ (منخفض) + ١٦.٥ (متوسط) + ١٢.٥ (مرتفع).
- توزيع النسب المئوية للدرجات الموزونة لمسنويات متغير الصعوبة إلى مجموع الدرجات الموزونة لمستويات المتغير = ١٢% منخفض + ٥٠% متوسط + ٣٨% مرتفع.

ويمكن اشتقاق الصيغ الخاصة بكل متغير من متغيرات البحث وعلى غرار ما سبق كما يلي:

- توزيع الدرجات الموزونة على متغير التمييز لمفردات الاختبار = ١٠.٥ (منخفض) + ١٤.٥ (متوسط) + ١١.٥ (مرتفع).
- توزيع النسب المئوية للدرجات الموزونة لمسنويات متغير التمييز= ٢٨% منخفض + ٤٠% متوسط + ٣٢% مرتفع.
- توزيع الدرجات الموزونة على متغير أقصى معلومات لمفردات الاختبار = ٢ (منخفض) + ١٢.٥ (متوسط) + ١٠ (مرتفع).
- توزيع النسب المئوية للدرجات الموزونة لمسنويات متغير أقصى معلومات = ٨% منخفض + ٥١% متوسط + ٤١% مرتفع.
- توزيع الدرجات الموزونة على متغير التخمين لمفردات الاختبار = ١١ (منخفض) + ٧ (متوسط) + ٣.٥ (مرتفع).
- توزيع النسب المئوية للدرجات الموزونة لمسنويات متغير التخمين = ٥١% منخفض + ٣٣% متوسط + ١٦% مرتفع.

- توزيع الدرجات الموزونة على متغير نوع مفردات الاختبار = ١٠(مقال) + ٩.٥(مزاوجة) + ٧(اختيار من متعدد) + ٢(صواب وخطأ) + ٥.٥(إجابة قصيرة) + ٤ (إكمال).
- توزيع النسب المئوية للدرجات الموزونة لمسنويات متغير نوع المفردة = ٢٦% مقال + ٢٥% مزاوجة + ١٨% اختيار من متعدد + ٥% صواب وخطأ + ١٤% إجابة قصيرة + ١٠% إكمال.
- توزيع الدرجات الموزونة على متغير المستوى المعرفي لمفردات الاختبار = ١.٥(تذكر) + ٣.٥(فهم) + ١٣(تطبيق)+ ٤.٥(عليا).
- توزيع النسب المئوية للدرجات الموزونة لمسنويات متغير المستوى المعرفي = ٧% (تذكر) + ١٥% (فهم) + ٥٨% (تطبيق)+ ٢٠% (عليا).
- توزيع توزيع الدرجات الموزونة على متغير تمثيل الاختبار لنطاق المتغير المقاس = ١٧.٥(كل النطاق)+ ١٢(ثلثي النطاق)+ ٣.٥(ثلث النطاق).
- توزيع النسب المئوية للدرجات الموزونة لمسنويات متغير تمثيل نطاق المتغير المقاس = ٥٣% (كل النطاق) + ٣٦% (ثلثي النطاق) + ١١% (ثلث النطاق).

ويقترح الباحث تقييم كفاءة الاختبار فيما يتعلق بتوزيع صعوبة فقراته كما يلي:

١- في حال اتساق نسب صعوبة أسئلة الاختبار مع النموذج السابق مع احتمال زيادة أو نقص أي نسبة من أي مستوى من المستويات الثلاث (المستوى السهل، المستوى المتوسط الصعوب، المستوى

المرتفع الصعوبة) زيادة أو نقص في حدود لاتتجاوز (± ٢%) يمنح الاختبار الدرجة (١٠) وهي أقصى درجة لكفاءة الاختبار من التوزيع حيث المثالي لصعوبة مفرداته.

٢- في حال تجاوز النسبة المقررة لكل مستوى من مستويات الصعوبة الثلاثة النسبة المحددة (± ٢%) (في ضوء نظرية البواقي المعيارية التي ينبغي ألا تتجاوز (± ٢%) عن ينتقض من الدرجة (١٠) قدر هذا التجاوز فعلى سبيل المثال إذا كان توزيع نسب مستويات مفردات الاختبار جاءت كمايلي:

١٠% منخفض + ٤٦% متوسط + ٤٠ % مرتفع.

يتضح من النموذج السابق لتوزيع نسب متغير صعوبة أسئلة الاختبار وفقا لمستوياته الثلاث (السهلة، المتوسطة، المرتفعة) يتضح أن:

- نسبة الأسئلة السهلة هي (١٠%) وبالتالي فقد انخفضت عن النسبة المثالية (١٢%) بمقدار (-٢%) وبالتالي لم تتجاوز النسبة المسموح بها، وكذلك نسبة الأسئلة الصعبة بلغت (٤٠%) وهي إزدادت بمقدار (+٢%) عن النسبة المثالية (٣٨%) فهي لم تتجاوز أيضاً النسبة المسموح بها.
- نسبة المفردات المتوسطة الصعوبة بلغت (٤٦%) وهي انخفضت عن النسبة المثالية بقدر (٤%) وبالتالي تجاوزت النسبة المسموح بها وهي (+٢%) وبالتالي فإن مقدار التجاوز (٤%- ٢%) ينبغي أن يخصم من الكفاءة النسبية للاختبار فيما يتعلق بمتغير الصعوبة. وبالتالي يجب أن يحصل الاختبار في تقييم كفاءته النسبية من حيث مستوى الصعوبة على الدرجة (٨).

١- يتبع نفس الإجراء مع النماذج الإحصائية التي تتعلق بالمتغيرات المتبقية (التمييز، أقصى معلومات، التخمين، نوع المفردات، تمثيل نطاق متغير القياس، المستوى المعرفي.

٢- المجموع الكلي للكفاءة النسبية للاختبار في ضوء النسب المثالية الواردة بالنماذج الإحصائية السابقة هي (١٠ × ٧ متغيرات) =٧٠.

٣- ويقترح أن الدرجة يجب أن تقييم الكفاءة النسبية للاختبار في ضوء المعايير المبينة بالجدول التالي:

**جدول (٧٠)**

**معايير الكفاءة النسبية للاختبار التحصيلي البنكي**

| مستوى الكفاءة النسبية | المؤشر الإحصائي للكفائة | مستوى الكفاءة النسبية | المؤشر الإحصائي للكفائة |
|---|---|---|---|
| متدني جدا | (١٠-٠) | فوق المتوسط | (٥٠-٤٠) |
| متدني | (٢٠-١٠) | مرتفع | (٦٠-٥٠) |
| تحت المتوسط | (٣٠-٢٠) | مرتفع جدا | (٧٠-٦٠) |
| متوسط | (٤٠-٣٠) | - | - |

**خاتمة:**

لقد أسفرت نتائج الدراسة الحالية عن مستويات محددة من المتغيرات الإحصائية والمعرفية يمكن أن تسهم في كفاءة الاختبار التحصيلي البنكي بأوزان مختلفة، وبالتالي يمكن الاعتماد عليها كمعايير في تقييم تلك الكفاءة، حيث تبين أن الأوزان الأكبر كانت لالمستوى المتوسط من صعوبة المفردات، وكذلك الحال مع المستوى المتوسط من التمييز، كما والمستوى المنخفض من التخمين، وكذلك المستوى المتوسط

والمرتفع من قيم المعلومات، وأيضا نسب تمثيل نطاق المتغير المقاس التي لا تقل عن (٦٦%).

كما أن الأسئلة المقالية وأسئلة المزاوجة والاختيار من متعدد كان لها أكبر الأوزان، ونفس الحال مع مستوى التطبيق والمستويات العليا بتصنيف بلوم للأهداف، كانت لها أكبر الأوزان في الاسهام بكفاءة الاختبار التحصيلي البنكي.

وبالتالي اشتقت الدراسة معايير لتقييم الكفاءة النسبية لهذا الاختبار في ضوء الأوزان المحددة لتلك المتغيرات بمستوياتها المختلفة كي تشكل نموذجا يمكن استخدامه في تقييم الكفاءة النسبية لأي اختبار تحصيلي بنكي يحتوي على مفردات محددة ومعلومة الخصائص بشكل مسبق.

**دراسات وبحوث مقترحة:**

في ضوء النتائج التي اسفرت عنها الدراسة يمكن اقتراح البحوث التالية:

- دراسة مقارنة بين الكفاءة النسبية للاختبارات التحصيلية البنكية في ضوء المعايير التي أسفرت عنها الدراسة الحالية، وتقديرات الكفاءة المقدرة من خلال الخبراء المختصون في مجال الاختبار.
- اشتقاق معايير لتقييم الكفاءة النسبية لاختبارات التحصيل في ضوء متغيري زمن الإجابة عن أسئلة الاختبار وعدد الأسئلة.
- دراسة العلاقة بين مستويات مختلفة من الكفاءة النسبية للاختبار التحصيلي البنكي (إعتمادا على المعايير المشتقة من الدراسة الحالية) وشكل توزيعات القدرة للأفراد المتلقين للاختبار.

## توصيات الدراسة:

توصي الدراسة في ضوء النتائج التي انتهت إليها بما يلي:

- إنشاء بنوك للأسئلة في مختلف المجالات التحصيلية لإتاحة وتوفير أرصدة من المفردات البنكية محددة ومعلومة المواصفات والخصائص الإحصائية كحل للاختبارات التحصيلية التقليدية التي تحتوي على مفردات لايتوافر عنها أية معلومات أو خصائص مسبقة، مما يجعل عملية تقويمها عملية ذاتية تفتقر إلى الموضوعية.
- سحب مفردات الاختبارات التحصيلية من أرصدة المفردات البنكية وبالتالي يمكن أن يتوافر لتلك الاختبارات صفات وملامح وخصائص واضحة ومحددة تمكن المقومين لها من استخدام المعايير التي انتهت إليها الدراسة الحالية بصورة علمية مدروسة وموضوعية.
- يمكن استخدام المعايير التي اشتقت من النتائج التي أسفرت عنها الدراسة الحالية في تقويم الاختبار التحصيلي التقليدي غير البنكي ولكن في ضوء تقديرات الخبراء في مجال الاختبار للخصائص والمؤشرات الخاصة بمفرداته والتي تتصل بمتغيرات الدراسة الحالية، كمتطلبات مسبقة وضرورية لاستخدام المعايير التي انتهت إليه الدراسة الحالية في تقييم كفاءته النسبية.

## المراجع:


- أحمد سليمان عودة (٢٠١٠). القياس والتقويم في العملية التدريسية. عمان: درا الأمل للنشر والتوزيع.
- أمينة محمد كاظم (١٩٨٨). دراسة نظرية نفقدية حول القياس الموضوعي للسلوك. الكويت: مؤسسة الكويت للتقدم العلمي.
- جواهر محمد الزيد (٢٠٠٧). فاعلية نظرية الاستجابة للمفردة في بناء الاختبارات (إعداد بنك أسئلة في مقرر علم النفس التربوي). رسالة دكتواره غير منشوره. جامعة الإمام محمد بن سعود الإسلامية. كلية العلوم الاجتماعية.
- رجاء محمود أبو علام (٢٠٠٩). التحليل الإحصائي للبيانات باستخدام برنامج spss: القاهرة: دار النشر للجامعات.
- سامي محمد ملحم (٢٠١١). القياس والتقويم في التربية وعلم النفس. عمان: دار المسيرة للنشر والتوزيع.
- صلاح الدين محمود علام (٢٠١٠). القياس والتقويم التربوي في العملية التدريسية. ط٣. عمان: دار المسيرة للنشر والتوزيع.
- عمر صالح ياسين (٢٠٠٤). الخصائص السيكومترية لاختبار محكي المرجع في الكيمياء. جامعة ام القرى. طلاب الصف الأول الثانوي العلمي مقدرة وفق النظريتين الكلاسيكة والحديثة للقياس. رسالة دكتوارة غير منشورة. كلية العلوم التربوية. جامعة عمان العربية.
- محمد شامل بهاء الدين فهمي (٢٠٠٩). الإحصاء بلامعاناة. الرياض: معهد الإدارة العامة. مركز البحوث.
- وفاء محمود نصار (٢٠١٠). القياس والتقويم في العملية التعليمية. ط١. الرياض: درا الزهراء.
- Aaron U., Irina K., David S.(2005). Authentic Learning:Balancing Bloom's Cognitive

**and Affective Domains, Teaching of Psychology, Vol. 32, No. 3,pp. 154-160.**

- **Altaf, Saim, Aslam, Muhammad.(2011). Bayesian Analysis Of The Davidson Model For Paired Comparison With Order Effect Using Noninformative Priors, Pakistan Journal of Statistics, 27 (2): pp.171-185.**
- **Anastasi, A.(1988): Psychological Testing, New York: Mac-Millan, 6th, ed.**
- **Azevedo,C.. (2009): Some Observations on the Identification and Interpretation of the 3PL IRT Model, Applied Psychological Measurement, 33(7): 89–114.**
-
- **Baird, Katherine (2011). Assessing Why Some Students Learn Math in High School: How Useful Are Student-Level Test Results?, Educational Policy, 25) 5(pp784-809.**
- **Begeny, J. C., Eckert, T. L., Montarello, S., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. School Psychology Quarterly, 23, 43–55.**
- **Best, W.J.,& Kahn, V.J. (2006). Research in education (11th ed.). Pearson Education Inc.**
- **Bowers, Alex J(2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. Educational Research & Evaluation, 17 (3), pp141-159.**

- **Burton, E. R. (2005). Multiple-choice and true- false tests: myths and misapprehensions. Assessment & Evaluation in Higher Education, 30(1), 65-72.**
- **Clifton SL ; Schriner CL(2010). Assessing the quality of multiple-choice test items. Nurse Educator (NURSE EDUC), 35(1): pp12-24.**
- **De Ayala,R.J.(2009).The Theory and Practice of Item Response Theory.NY.the Guilford press.**
- **De Champlain,A(2010): A primer on classical test theory and item response theory for assessments in medical education,Blackwell Publishing Ltd, Medical Education ; 44(1):109–117.**
- **de la Torre, Jimmy, Hao Song, Yuan Hong.(2011). A Comparison of Four Methods of IRT Subscoring. Applied Psychological Measurement, Vol. 35 Issue 4, p296-316.**
- **Eckert, T. L., Dunn, E. K., Codding, R. S., Begeny, J. C., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. Psychology in the Schools, 43, 247–265.**
- **Edward, H. (2010). Interpretation of the Three-Parameter Testlet Response Model and Information Function. Applied Psychological Measurement, 34 (7): pp467-482.**
- **Hicks, T. ; Lewis, L. ; Munn, G. ; Jordon, E. ; Charles, K.(2010). An Assessment of Teacher Education Students' Perceptions and Satisfaction of Their Learning**

**Experiences in a Summer Pilot Program, College Quarterly, v13 n1. pp 19.**

- **Huang, J.(2010) Applying Item Response Theory in Language Test Item Bank Building, Modern Language Journal;, Vol. (2),p p.374-385.**
- **Hudson, S., Lewis, T., Stichter, P.(2011) Putting Quality Indicators to the Test: An Examination of 20 Years of Research, Applied Psychological Measurement, 20 (7): pp115-127.**
- **Hurwitz, J. T., Elliott, S. N., & Braden, J. P. (2007). The influence of test familiarity and student disability status upon teachers' judgments of students' test performance.,School Psychology Quarterly, 22, 115–144.**
- **Hurwitz, J. T., Elliott, S. N., & Braden, J. P. (2007). The influence of test familiarity and student disability status upon teachers' judgments of students' test performance. School Psychology Quarterly, 22, 115–144.**
- **James A. Hills (2011).Better Teaching And With Deming and Bloom,, Quality Progress, 37 (3),pp.54-64.**
- **Jandaghi, G.,& Shateria, F. (2008). Rate of validity, reliability and difficulty indices for teacher-designed exam questions in first year high school. International Journal of Human Sciences,5(2).**
- **Jarle Moen, Martin Tjelta (2010). Grading Standards, Student Ability and Errors in College Admission, Scandinavian Journal of Educational Research, 54(3)pp.221–237.**

- **John C..,Krouse, G., Courtney M. (2011) Teacher Judgments of Students' Reading Abilities Across a Continuum of Rating Methods and Achievement Measures, School Psychology Review, 40(1), pp. 23–38**
- **Katie. A.(2011). Tailoring Future Testing. Education Week, 30 (25), pp.35-39.**
- **Kingsbury & Zara.(2002). Item Selection in Computerized Adaptive Testing: Should More Discriminating Items be Used First? Journal of Educational Measurement, 33 (2): pp221-255.**
- **Kucuk, F., Walters, J. (2009). ELT Journal: English Language Teachers Journal,63 (4), pp332-341.**
- **Lehman, A. ; Woodbury, M ; Shechtman, O.(2011). Development of an item bank for a computerised adaptive test of upper-extremity function. Disability & Rehabilitation (DISABIL REHABIL), 33(21): 290-214.**
- **Lihua Yao, Schwarz, D.(2006). A Multidimensional Partial Credit Model With Associated Item and Test Statistics: An Application to Mixed-Format Tests. Applied Psychological Measurement, 30 (6),pp.469-492.**
- **Lihua Yao.(2010).Reporting Valid and Reliable Overall Scores and Domain Scores. Journal of Educational Measurement, 47 (3): pp339-360**
- **Lihua Yao.(2011). Multidimensional Linking for Domain Scores and Overall Scores for Nonequivalent Groups. Applied Psychological Measurement, 33 (2): pp40-68.**

- **Linacre J.M., (2005). A User's Guide to W I N S T E P S & M I N I S T E P, Chicago:Scientific Software Inc.**
- **Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.**
- **Magis, David.(2011). A Test-Length Correction to the Estimation of Extreme Proficiency Levels, Applied Psychological Measurement, Vol. 35 Issue 2, p91-109.**
- **Marinagi, Catherine C, Kaburlasos, Vassilis G.(2010). Bayesian Decision Theory for Multi-Category Adaptive Testing, Educational & Psychological Measurement, 20(1):p p37-59.**
- **Meisinger, E. B., Bradley, B. A., Schwanenflugel, P. J.,Kuhn, M. R., & Morris, R. D. (2009). Myth and reality of the word caller: The relation between teacher nominations and prevalence among elementary school children.School Psychology Quarterly, 24, 147–159.**
- **Newhouse, C. Paul(2011). Using IT to Assess IT: Towards Greater Authenticity in Summative Performance Assessment, Computers & Education, v56 n2, p p 388-402.**
- **Nilay T. Bumen (2007). Effects OF The Original Versus Revised BLOOM'S Taxonomy ON Lesson Planning Skills and test design: A Turkish Study Among Pre-Service Teachers.**
- **Norbert S. and Janice W. (2011). The Word Associates Format: Validation evidence, Language Testing,28(1),pp.105–126.**
- **Nunnally, J.C., & Bernstein, I.H. (1994) (3rd ed.). Psychometric theory. New York: McGraw Hill.**

- **Paul, J.(1991). Bayesian interpretation of test reliability. Journal Of Educational And Psychological Measurement, 51, pp.627-635.**
- **Richard F. Burton (2001). Do Item-discrimination Indices Really Help Us to Improve Our Tests? Assessment & Evaluation in Higher Education, 26(3),pp.213-220.**
- **Richard F. Burton (2004). Multiple choice and true/false tests:reliability measures and some implications of negative marking, Assessment & Evaluation in Higher Education, 2(5),pp.585-595.**
- **Roach, Andrew T, Beddow, Peter A, Kurz, Alexander,(2010) Incorporating Student Input in Developing Alternate Assessments Based on Modified Academic Achievement Standards, Research Papers in Education; 77 (1), pp90-112.**
- **Rose,B. Bjorner, J. Becker, F. Friesc,.E. (2008). Evaluation Of A Preliminary Physical Function Item Bank Supported The Expected Advantages Of The Patient-Reported Outcomes Measurement Information System (Promis), Journal of Clinical Epidemiology,61(4): pp17-33.**
- **Ryan J., Peter A., Elizabeth C.(2010) What Do Alternate Assessments of Alternate Academic Achievement Standards Measured A Multitrait—Multimethod Analysis,Exceptional Children, 76, (4), pp. 457-474.**
- **Schmitt, T. A.(2010). A Monte Carlo Simulation Investigating the Validity and Reliability of Ability Estimation in Item Response Theory with Speeded**

**Computer Adaptive Tests. International Journal of Testing, 10 (3):p p230-261.**

- **Schurmeier, Kimberly D., Atwood, Charles H, Shepler, Carrie G., Lautenschlager, Gary J.(2010). Using Item Response Theory To Assess Changes in Student Performance Based on Changes in Question type. Journal of Chemical Education, 87 (11), pp1268-1272**
- **Schuwirth, L., van der Vleuten, C. (2012).Programmatic assessment and Kane's validity perspective. Medical Education, 46 (1), pp38-48.**
- **Siang C. & Fritz D.(2006). How Big Is Big Enough? Sample Size Requirements for CAST Item Parameter Estimation, Applied Psychological Measurement, 19(3):pp 241–255.**
- **Stringer, Neil (2008). Aptitude tests versus school exams as selection tools for higher education and the case for assessing educational achievement in context.. 23 (1) pp53-68.**
- **Swanson D,. Holtzman K., Butler A.(2010). Cumulative achievement testing: Progress testing in reverse, Medical teaching, 2010; 32pp: 516–520**
- **Swiatek, Mary Ann (2005). Gifted Students' Self-Perceptions of Ability in Specific Subject Domains: Factor Structure and Relationship With Above-Level Test Scores. Roeper Review, 27(2), pp.104-109**
- **Tasdemir, Mehmet (2010). A Comparison of Multiple-Choice Tests and True-False, Matching Tests Used in Evaluating Student Progress, Journal of**

**Instructional Psychology, 37(3),pp.258-266.**

- **Tavakol, Mohsen, Dennick, Reg.(2011) Post-examination analysis of objective tests. Medical Teacher; 33(6),pp447-458.**
- **Tie Liang, W., Craig S.(2009). A Model Fit Statistic for Generalized Partial Credit Model, Educational & Psychological Measurement, 69 (6) pp.913-928.**
- **Tyrone B.,David L. (2011) evaluating and improving student achievement in business programs: the effective use of standardized assessment tests, journal of education for business,30(4),pp.24-44.**
- **Zimowski, M.,Muraki,E.,. Mislevy,R., Bock,R.(2007). BILOG-MG 3 for Windows, Chicago:Scientific Software Inc.**