

استخدام نموذج راش في إعداد وتقنين اختبار في اللغة العربية لقبول طلاب الدراسات العليا بكليات العلوم الإنسانية بجامعة الكويت

د. فوزية عباس هادي أ. د. صلاح أحمد مراد

كلية التربية - جامعة الكويت

ملخص

هدفت هذه الدراسة إلى إعداد وتقنين اختبار في اللغة العربية لقبول الطلاب ببرنامج الدراسات العليا بكليات العلوم الإنسانية بجامعة الكويت. وتمت مراجعة الأدبيات والاختبارات المتوافرة في هذا المجال، وبناء عليه تم تحديد ثلاثة محاور أساسية لیتضمنها الاختبار وهي: القدرة على الفهم، والتعبير، والتفكير. وتم إعداد ١٣٠ بنداً وتدريبها على ٨٦ طالباً وطالبة بالدراسات العليا وأسفر التجريب عن حذف ٣٠ بنداً وتعديل ١٦ بنداً. وتم تطبيق الصورة النهائية على عينة حجمها ٣٣٥، وتم تحليل البيانات بالطريقة العادية وباستخدام نموذج راش، وأسفرت نتائج الطريقة العادية عن حذف ١٧ بنداً وبلغ معامل كرونباك ألفا ٠,٨٦. وتراوحت معاملات الاتساق الداخلي بين ٠,٤٦، ٠,٨٨، كما ارتبطت درجات الاختبار بالمعدل التراكمي ارتباطاً دالاً وجميعها دالة عند ٠,٠٠١.

أما التحليل باستخدام نموذج راش في تدریج بنود الاختبار (١٠٠ بند) فقد نتج عنه حذف ٢٦ فرداً غير ملائمين للنموذج وحذف ١٩ بنداً غير ملائمة للنموذج، وأسفر التحليل عن اختبار جيد يتمتع بثبات ٠,٨١، لبنود الاختبار، ٠,٩٩، لقدرات الأفراد، ومدى القدرة يتراوح بين $2 \pm$ لوجيت بمتوسط = ٠,٥٠، لوجيت وهو مستوى جيد.

وتم إعداد جداول معايير للاختبار باستخدام اللوجيت والدرجة المعدلة والدرجة الجامعية (مشابهة لدرجات التوفل) كما تم تحديد درجة القطع لاختيار الطلاب لبرنامج الماجستير في كليات العلوم الإنسانية.

مقدمة

يُعد الاستثمار البشري من أفضل الاستثمارات في دول العالم التي تود تعظيم الاستفادة من إمكاناتها البشرية، وفي سبيل ذلك فهي تبذل الجهد وتضع الخطط والبرامج التي تتيح الفرص لمواطنيها لتنمية قدراتهم إلى أقصى درجة ممكنة وفقاً لاستعداداتهم ورغباتهم. وهي بذلك أيضاً تستفيد من قدراتهم وابداعاتهم في أداء أعمالهم وتنمية مجتمعاتهم. كما أنها تُريد من فاعليتهم في القيام بأدوارهم اللازمة لتحقيق الخطط التنموية ومواكبة التقدم العالمي.

وفي هذا الشأن، فإن دول العالم المتقدم تولي اهتماماً بالغاً بتعليم أبنائها تعليماً جيداً، واكسابهم المعارف والمهارات التي تُلبّي احتياجاتهم وتُثمي قدراتهم ليكونوا عناصراً للتقدم والتطور. ومن هذا المنطلق فإن التعليم في مثل هذه الدول يُركز على الاستخدام الأمثل لقدرات الأبناء إلى أقصى درجة ممكنة، بهدف تنميتها وتشجيعها وتفعيلها وحصد مبتكراتها واختراعاتها.

لذا نجد أن الجامعات في هذه الدول تضع شروطاً كثيرة عند انتقاء الطلبة للدراسات العليا. ومن بين هذه الشروط المعدل التراكمي واجتياز اختبارات قدرات واستعدادات مقننة للقبول تختلف باختلاف التخصص والبرنامج الدراسي، وتضع أيضاً من الإجراءات والشروط ما يضمن انتقاء أفضل المتقدمين.

مشكلة البحث

تتضمن لائحة كلية الدراسات العليا بجامعة الكويت شروطاً محددة لكل برنامج يخضع لها الطلبة الراغبون في التقديم من أجل مواصلة دراساتهم العليا في البرنامج المناسب للدرجة الجامعية الأولى، ثم تتم المفاضلة بين هؤلاء المتقدمين وفق معدلاتهم التراكمية واجتياز المقابلة الشخصية أو الاختبار الذي تعقده لجنة البرنامج للطلبة المتقدمين، وجميع هذه الإجراءات مناسبة وتهدف إلى مراعاة الحياد والموضوعية في انتقاء أفضل المتقدمين.

ومن شروط القبول ببرامج كلية الدراسات العليا بجامعة الكويت أيضاً اجتياز اختبار التوفل TOEFL الذي يُعد عقبة كبرى لبعض الطلبة المتقدمين. ولوحظ أن كثيراً من طلبة الدراسات العليا يفشلون في الحصول على الدرجة المطلوبة للإلتحاق ببرنامجهن الدراسي لضعفهن في اللغة الإنجليزية، وعلى الرغم من أن بعض الطلاب يرون أن اختبار التوفل TOEFL غير مناسب لأن دراستهن باللغة العربية ويحاولون إيجاد مبررات لتخفيض درجات التوفل TOEFL المطلوبة للدراسات العليا، إلا أنه لوحظ ضعف كبير لدي طلبة الدراسات العليا في اللغة العربية أيضاً. حيث أشار عدد من الأساتذة أن كثيراً من الدارسين ببرامج الدراسات العليا يجدون صعوبة في كتابة تقاريرهم البحثية بأسلوب مناسب وجيد، ويواجهون صعوبة أكبر في كتابة الأطروحة اللازمة للتخرج، الأمر الذي يؤدي بعدد كبير من الطلبة إلى اختيار مسار اللأطروحة في برامج الماجستير. وفي هذا دلالة واضحة على ضعف في اللغة العربية أو في أساليب التعبير اللغوي أيضاً.

وأدى هذا بالمسؤولين في كلية الدراسات العليا في عام ٢٠٠٨ إلى التفكير في وضع الحلول المناسبة لهذه المشكلات، ومنها انبثقت فكرة اعداد اختبار للقبول ببرامج الدراسات العليا في اللغة العربية، وأن تكون البداية مع طلبة برامج العلوم الإنسانية (في كليات التربية والآداب والشرعية والعلوم الاجتماعية والعلوم الإدارية والحقوق).

وجاء التكليف بإعداد هذا الاختبار من عميد كلية الدراسات العليا للباحثين^(١) (من قسم علم النفس التربوي) لدراسة المقترح وبحث إمكانية إنجاز المطلوب خلال عام دراسي واحد، وتم الاستعانة بأربعة^(٢) من أساتذة اللغة

(١) خطاب عميد كلية الدراسات العليا المرقم (س/ع/٢٦٢) بتاريخ ٢٠٠٨/٦/٣ بالموافقة للباحثين القيام بدراسة علمية ونشرها بالمجلات العلمية.

(٢) نشكر أعضاء اللجنة: أ. د. سهام الفريح، أ. د. تركي المغيض، أ. د. عبدالفتاح الحموز، د. محمد نشمي العنزي، على جهودهم في إعداد بنود الاختبار.

العربية بكلية الآداب للمساعدة في إعداد الاختبار وفق المواصفات المحددة لاختبار مشابه لاختبار التوفل TOEFL، وتم إجراء لقاءات عدة وبُذلت جهود مكثفة لأداء المطلوب خلال الفترة الزمنية المحددة.

اهتمت الدول المتقدمة بوضع معايير وضوابط لانتقاء الطلبة الراغبين بالالتحاق بالدراسة الجامعية، وبدأت مؤسسة الاختبارات التربوية في الولايات المتحدة الأمريكية (ETS) (Educational Testing Services) في إعداد اختبارات للقبول منها اختبار (SAT) (Scholastic Aptitude Test) الذي بدأ استخدامه عام ١٩٢٦، وهو اختبار استعداد للدراسة الجامعية يؤديه الطلبة الراغبين في الالتحاق بالجامعة لمرحلة البكالوريوس. ويتكون الاختبار من جزئين: لفظي، وكمي، وهو يقيس القدرة على القراءة النقدية ويتضمن نوعين من الأسئلة: أسئلة إكمال الجمل، وأسئلة تفسيرية (وهي أسئلة تأتي بعد مقدمة مختصرة من المعلومات). وقد صُممت أسئلة إكمال الجمل لقياس معلومات الطالب في فهم معاني الكلمات وتكوين الجمل. بينما الأسئلة التفسيرية تعتمد على مقدمة تختلف في الطول والنمط والنوع وتهتم بفهم السياق والاستدلال منه (Kirkup, Schagen, Wheeler, Morrison & Whetton, 2007). وقد أثبتت البحوث فعاليتها في التنبؤ بالأداء الأكاديمي بالجامعة على الرغم من أن معدل الثانوي أفضل من درجات SAT في التنبؤ بالنجاح الأكاديمي (Geiser, 2008-a)، لكن معدل الثانوي جاء نتيجة تأدية اختبارات مختلفة بينما درجات SAT تعتمد على صور متكافئة لاختبار واحد. وقد أُجري تعديل على اختبار SAT عام ٢٠٠٥ بإضافة اختبار فرعي للكتابة، وقد ثبتت فعالية هذا الاختبار الجديد في التنبؤ بالنجاح الجامعي أفضل من الاختبار القديم (Geiser, 2008-b).

وقد أشار تانر (Tanner, 2003) إلى أن الصدق التنبؤي لدرجات SAT في العام الجامعي الأول هو الذي أعطى قيمة للاختبار بينما علاقته مع GPA تتراوح بين ٠,٣ - ٠,٥ بمعنى أنه يُفسر ٢٥٪ من تباين التحصيل. لكن رأي (Ray, 2008) توصل إلى معاملات ارتباط مرتفعة بين درجات SAT ومعدل

الثانوي بلغت ٠،٧٦، ٠،٨٨، ٠،٧٣، ٠،٧٠ للاختبار الكمي، ٠،٧٣، ٠،٧٠ للاختبار اللفظي. وقد ذكر (Menson,Patelis & Doyle, 2009) أن البحوث على مدى عقود وجدت أن درجات SAT أفضل منبئ للتحصيل الجامعي في العام الأول، كما أشارت بعض البحوث إلى أن دمج درجات SAT مع معدل الثانوي يُعد أفضل منبئ بالأداء الجامعي.

وقد أجرى رجاء أبوعلام وفتحي الديب (١٩٨٥) دراسة للتعرف على مدى قدرة درجات الثانوية العامة على التنبؤ بالأداء الأكاديمي لطلبة جامعة الكويت، وبلغت العينة ١١٧٣ طالبا وطالبة (٧٠٢ أدبي، ٤٧١ علمي) وتم الحصول على المعدلات التراكمية للطلبة في ستة فصول دراسية متتالية من ١٩٧٩ وحتى ١٩٨١. وتوصلا إلى وجود معاملات ارتباط دالة بين درجات الثانوية العامة والمعدل التراكمي لكل فصل دراسي. وتراوحت معاملات الارتباط للمعدل العام بين ٢٩٦ - ٣٩٦. للأدبي، ٤٦٦ - ٥٨٣. للعلمي، في حين تراوحت معاملات الارتباط بالمعدل التراكمي للتخصص بين ١١٧ - ٢٤٦. للأدبي، ١٣٣ - ٣٠٥. للعلمي. وتدل هذه الارتباطات على إمكانية التنبؤ بالأداء الأكاديمي باستخدام درجات الثانوية العامة مع أن هذه القدرة التنبؤية ضعيفة.

كما أجرت نادية شريف (١٩٨٨) دراسة مقارنة للقيمة التنبؤية لدرجات الثانوية العامة لنظامي المقررات والتقليدي في النجاح بالمرحلة الجامعية. واختارت عينة عشوائية من طلبة جامعة الكويت حجمها (٣٦٩٩)، وبلغ حجم العينة ٥٦٢ (٢٥٦ ذكور، ١٣٠ إناث) منهم ١٧٨ من نظام المقررات، ٣٨٤ من النظام التقليدي. وتم حساب معاملات الارتباط بين درجات الثانوي (مقررات، تقليدي) مع المعدل التراكمي للطلبة في آخر فصل دراسي بالجامعة، وتراوحت معاملات الارتباط بين درجات الثانوي والمعدل التراكمي بين ٥٢ - ٦٩، لنظام المقررات وبين ٥٤ - ٥٩، للنظام التقليدي ومعامل ارتباط واحد غير دال للطلاب الذكور أدبي. وتدل هذه المعاملات على تحسن القدرة التنبؤية لدرجات الثانوية العامة في التنبؤ بالأداء الجامعي.

وأجرى عبدالناصر أحمد (٢٠٠٧) دراسة لاعداد مقياس للقبول بالجامعات المصرية يقيس الاستعداد اللغوي والرياضي واللغة الانجليزية والميول نحو الدراسة الجامعية، وطبق المقياس على ٥٥ طالب وطالبة بجامعة الفيوم لتدريج المقياس (Calibration). ثم استخدم المقياس مع ٢١٣ طالب وطالبة بكلية التربية بجامعة الفيوم، وتوصلت الدراسة إلى أن مقياس القبول يتنبأ بـ ٤٢٪ من الأداء الأكاديمي الجامعي وبإضافة درجات الثانوية العامة أمكن التنبؤ بـ ٥٤،٢٪ من الأداء الأكاديمي.

وتوصل (Bridgeman, McCamley-Jenkins & Ervin, 2000) في دراسة للتنبؤ بالنجاح الأكاديمي باستخدام عينة قوامها ٤٨٠٠٠ طالب وطالبة من ٢٣ كلية إلى أن التنبؤ بالنجاح الأكاديمي باستخدام اختبار SAT تراوح بين ١٩،٠ - ٣،٠ للاختبار اللفظي، و ١٩،٠ - ٣٤،٠ للاختبار الكمي. ووجد عبدالله الصريفي (٢٠٠١) أن معدل الثانوية العامة يرتبط بأداء طلاب الكليات العملية (٥٤٥،٠) والطالبات (٤٦٧،٠) وأن قيمة التنبؤ باستخدام المعدل العام ٥٩١،٠ واستخدمت فوقيه السيد (١٩٩٥) عدة مقاييس لبحث القدرة التنبؤية لنجاح طلاب كلية الهندسة وكانت عينة الدراسة ٨٠٠ طالب وطالبة من كليات الهندسة بأربع جامعات مصرية، وتوصلت الدراسة إلى أن الإستعداد الميكانيكي والذكاء العام كانا أعلى المنبئات.

وحيث إن القدرات اللازمة للنجاح الجامعي تختلف عن القدرات التي تتنبأ بالنجاح في الدراسات العليا، فقد أعدت ETS اختباراً آخر لقبول طلبة الدراسات العليا وهو Graduate Record Examination (GRE). ويختلف GRE عن SAT في أنه يقيس: القدرة على التحليل، وتقويم المعلومات المقدمة، وتركيب الأفكار، والاستدلال وحل المشكلات (Tanner, 2003). والقدرة التنبؤية لاختبار GRE أقل من اختبار SAT فقد وجد موريسون وموريسون (Morrison & Morrison, 1995) أن علاقة التحصيل مع درجات GRE اللفظي ٢٨،٠ والكمي ٢٢،٠، وقد يرجع ذلك إلى أن قدرات طلبة الدراسات العليا أقل تباينا من طلبة الجامعة.

وأجرى تانر (Tanner, 2003) دراسة باستخدام ثلاث اختبارات للقبول بجامعة كاليفورنيا وهي: SAT، GRE، واختبار آخر يقيس القدرة على فهم المعلومات المقروءة وفهم الجداول والأشكال والرياضيات ومهارة الكتابة وتوصل (Tanner, 2003) إلى أن العلاقة بين درجات SAT اللفظي ودرجات GRE اللفظي بلغت ٨٢,٠ كما ارتبطت درجات جميع الاختبارات الفرعية للاختبارات الثلاث المستخدمة ارتباطات دالة، وقد جمعها التحليل العاملي في عامل واحد يُفسر ٦٢٪ من التباين. وهذا يدل على تقارب الاختبارات المستخدمة للقبول في جامعة كاليفورنيا مما يدعم فكرة استخدام الاختبار الذي أعدته كاليفورنيا بديلاً لاختباري SAT و GRE.

وقام باورس ورفقاؤه (Powers, Swinton & Carlson, 1977) بدراسة عاملية على اختبار GRE حيث استخدم صورتين متكافئتين من الاختبار يحتوي كل منهما على ١١٠ سؤالاً (٥٥ لفظي، ٥٥ كمي)، حيث يقيس الاختبار اللفظي: فهم العلاقات بين الأفكار، وفهم معاني الكلمات، وتركيب الجمل، وفهم السياق في العبارات والتعبير المنطقي، وفهم المقروء بتقديم فقرات يليها أسئلة تدور حول الفكرة الرئيسية والاستنتاج والتطبيق والتقويم المنطقي وأسلوب التعبير. أما الاختبار الكمي فهو يقيس: جوانب كمية (حساب، جبر، هندسة، جداول، خرائط، أشكال) عن طريق إعطاء مقدمة يليها عدة أسئلة اختيار من متعدد. وطبق الاختبار على ٢٠٠٠ طالب وطالبة بواقع ١٠٠٠ لكل صورة، وأجري تحليل عاملي استكشافي لدرجات الصورة (أ) توصل إلى وجود ثلاثة عوامل (عاملان لفظيان، وعامل كمي) تُفسر ٧٦,٥٪ من التباين، وتوصل التحليل العاملي لدرجات الصور (ب) إلى وجود ثلاثة عوامل أيضاً تُفسر ٧٢٪ من التباين.

ودرس إيدي (Eddy, 1981) أثر تعلم لغة أجنبية لفترة أطول على الأداء على اختبار SAT ووجد أن الدرجات المرتفعة في اللغة الأجنبية يصاحبها درجات مرتفعة على اختبار SAT وخاصة الجزء الخاص بالقراءة والمفردات. كما اهتمت دراسة كوبر (Cooper, 1987) بأثر دراسة لغة أجنبية على درجات اختبار SAT ووجد أنها تؤدي إلى ارتفاع درجات SAT اللفظية.

وأعاد كوبر وآخرون نفس الدراسة لبحث أثر تعلم لغة أجنبية على الأداء الأكاديمي (Cooper, Yanasky, Wisenbaker, Jahnar, Webb & Wilbur, 2008) حيث أجريت الدراسة على ١٧٧٨ طالباً وطالبة في جنوب الولايات المتحدة، ووجدوا أنه بعد عزل أثر القدرة اللغوية فإن الطلبة الذين درسوا لغة ثانية ولفترة طويلة حصلوا على درجات مرتفعة في اختبار SAT اللفظي. وذلك لأن دراسة اللغة تؤدي إلى فهم معاني مفردات جديدة وفهم السياق واستخدام التعبيرات الجديدة، وتنمية مهارات التفكير الناقد التي يستخدمها الطالب (شعورياً أو لا شعورياً) في الأداء والمهام المكلف بها.

ونظراً لتزايد عدد المتقدمين لبرامج الدراسات العليا فإن أسلوب الاختيار قد يلجأ إلى استخدام محكات معينة تؤدي إلى استبعاد الجادين والتميزين وقبول عدد من ضعيفي المستوى مما يؤثر على المخرجات التعليمية، وقد يلجأ المسؤولون إلى تعديل في اللوائح وقبول عدد أكبر من المتقدمين، مما يؤثر على الأداء في البرامج ومن ثم مستوى مخرجاتها. لكنه من المفضل وضع محكات موضوعية وأكثر دقة تمكن من إنتقاء أفضل المتقدمين، وذلك من خلال ترتيبهم على هذه المحكات الموضوعية ترتيباً تنازلياً واختيار أفضلهم.

تناولت البحوث السابقة إجراءات القبول في برامج الدراسات العليا، التي تعتمد على ترتيب المتقدمين وفقاً لمعدلاتهم التراكمية GPA والمقابلات الشخصية، لكن من المفضل وجود معيار آخر موضوعي بديلاً عن المقابلات. وتستخدم الجامعات الأمريكية درجات GRE مع GPA (دون اهتمام بالمقابلات الشخصية)، إضافة إلى درجات التوفل TOEFL للطلبة الأجانب. وتوصلت دراسة باسترك (Basturk, 1999) إلى وجود علاقات دالة بين درجات GRE و GPA بلغت ٠,٤٣ على عينة حجمها ٥٢٨، وعلاقة ٠,٤٨ على عينة حجمها ٦١٩ من طلاب الماجستير، وارتبطت درجات تقويم أعضاء هيئة التدريس للطلبة مع درجات GRE ارتباطاً ضعيفاً (٠,٢٩). وبصفة عامة فإن مثل هذه العلاقات تُعد متحيزة بسبب حدود العينات المستخدمة وذلك لعدم قبول الحاصلين على

درجات منخفضة في الـ GRE. فإذا كان أفراد العينة مختلفين تزداد القدرة التنبؤية لدرجات الاختبار. كما أن معظم طلبة الدراسات العليا يحصلون على تقديرات A و B مما يقلل من تباين التحصيل وبالتالي يقلل إمكانية التنبؤ. فإذا كان المنبئ أو المحك متغيرات غير متباينة أو منخفضة التباين فإن ذلك يقلل من القدرة التنبؤية.

وتشترط كلية الدراسات العليا بجامعة الكويت لقبول الطلبة في برامجها الحصول على معدل تراكمي ثلاث نقاط أو أعلى، والحصول على ٤٥٠ درجة في اختبار TOFEL، إضافة إلى اجتياز المقابلة الشخصية وربما اجتياز الاختبارات التي تضعها الأقسام العلمية. وقد أدى هذا إلى عدم علاج ضعف أداء طلبة الدراسات العليا كما جاء في ملاحظات أعضاء هيئة التدريس، وشكواهم من تدني مستوى أداء الطلبة في كتابة تقارير البحوث والأطروحات العلمية.

وبناء على ما سبق كان التفكير في وضع معيار موضوعي آخر لقبول طلبة الدراسات العليا في تخصصات العلوم الإنسانية بجامعة الكويت، وهو إعداد اختبار مقنن في اللغة العربية يقيس القدرات اللازمة التي تمكن طالب الدراسات العليا من فهم المادة العلمية وكتابة التقارير البحثية المطلوبة منهم وإعداد الأطروحات العلمية.

نموذج راش Rasch Model

اهتم عالم الرياضيات الدنماركي جورج راش George Rasch بإرساء دعائم الموضوعية في سلوك الإنسان وقياس الظواهر السلوكية، حيث تبنى الأساس الإحصائي في تفسير الأداء. وافترض أن إجابة أي فرد على مفردة معينة تتحدد من مستوى قدرته، وهي التي تحدد احتمال حدوث الإجابة الصحيحة على مفردات الاختبار المتدرجة على متصل السمة المقاسة (Write & Stone, 1979)؛ مراد وسليمان، (٢٠٠٥).

ومن المفترض أن زيادة قدرة الفرد عن صعوبة المفردة يؤدي إلى زيادة

احتمال حدوث الاستجابة الصحيحة. لكن العلاقة بين قدرة الفرد وصعوبة المفردة ليست علاقته قاطعة، فقد تكون القدرة مرتفعة لكنه يجيب إجابة خاطئة وقد تزداد صعوبة المفردة على قدرة الفرد لكنه يجيب إجابة صحيحة. ولذلك توصل راش إلى أن الفرق بين القدرة والصعوبة هو الذي يحدد احتمال الإجابة الصحيحة (كاظم، ١٩٨٨). ويكون هذا الفرق مساوياً لو غاريتم (احتمال النجاح على المفردة ÷ احتمال الخطأ على المفردة).

ويفترض نموذج راش: (١) أحادية القياس وتُعني أن مفردات الاختبار تقيس سمة واحدة وتختلف فقط في مستوى الصعوبة، (٢) استقلالية القياس وتُعني عدم تأثر استجابة الفرد لمفردة باستجاباته على المفردات الأخرى، أو على قدرة الأفراد، (٣) توازي المنحنيات المميزة للمفردات وتُعني تساوي قوة تمييز المفردات للأفراد مختلفي القدرة. وهذا افتراض يدل على موضوعية القياس وهي تعني استقلال قدرة الفرد عن المفردة المستخدمة Item - Free، واستقلال صعوبة المفردة عن الأفراد Person - Free (Wright, Mead & Bell, 1980).

وتختلف نماذج السمات الكامنة باختلاف شكل المنحنى المميز للمفردة وبالتالي اختلاف صورة الدالة الرياضية (علام، ١٩٨٦).

استخدام نموذج راش في بناء الاختبارات

استخدام نموذج راش في بناء اختبار استعداد يعني توافر الموضوعية في قياس القدرة، ويدل هذا على توافر شرطي الصدق والثبات لتقديرات صعوبة البنود وقدرات الأفراد. وتتحدد مطالب الموضوعية في تعريف بنود الاختبار لمتغير واحد، واستقلال تقديرات الأفراد عن البنود، واستقلال صعوبة البنود عن أفراد العينة (كاظم، ١٩٨٨).

ويقوم البرنامج المستخدم لتحليل بيانات الاختبار (WINSTEPS) بحذف الأفراد غير الملائمين للنموذج وهم غير الصادقين في استجاباتهم على مفردات الاختبار وفقاً لقدراتهم. كما يتم حذف البنود غير الملائمة وهي التي لا تتسق

مع البنود الأخرى على المتغير موضع القياس. ويتم تدريج بنود الاختبار مستقلة عن الأفراد، وتقدير أداء الفرد مستقلاً عن البنود. وينتج عن ذلك تدريج بنود الاختبار بوحدات متساوية تسمى اللوجيت Logit (كاظم، ١٩٨٨، الشافعي، ١٩٩٦).

دراسات سابقة استخدمت نموذج راش في بناء الاختبارات

أجرى هيننج (Hennig, 1984) دراسة لمقارنة نتائج استخدام نموذج راش بالأسلوب التقليدي في تحليل اختبار في اللغة الإنجليزية وكانت العينة ١٠٨ أفراد يدرسون مقرر في اللغة الإنجليزية كلغة ثانية، وطبق عليهم اختباراً مكون من ٤٨ مفردة من نوع الاختيار من متعدد ولكل منها أربعة بدائل. وتوصلت الدراسة إلى أن نموذج راش يتيح معلومات أفضل من الأسلوب التقليدي وقد ظهر ذلك في ارتفاع معامل الثبات بعد حذف المفردات غير الملائمة، كما يوفر النموذج قيمة للخطأ المعياري لصعوبة كل مفردة ولقدرة كل فرد بعكس الطرائق التقليدية التي توفر قيمة واحدة للخطأ المعياري للاختبار.

وأجرى فوربس (Forbes, 1986) دراسة لبحث فعالية استخدام نموذج راش في بناء الاختبارات التحصيلية في اللغة الإنجليزية والعلوم والرياضيات، وكان حجم العينة ٨٠٠٠ طالب وطالبة، وتوصل من دراسته إلى أنه يمكن إعداد اختبارات تحصيلية أكثر جودة باستخدام نموذج راش. ودرست أمينة كاظم (١٩٨٨) استخدام نموذج راش في بناء اختبار تحصيلي في مقرر علم النفس بجامعة الكويت وبلغ حجم العينة ٤١٨ طالباً وطالبة، وتكون الاختبار من ١٠٠ مفردة من نوع الاختيار من متعدد ولكل منها أربعة بدائل. وتم حذف ٢٣ فرداً غير ملائمين وخمس مفردات غير ملائمة لأسس القياس الموضوعي. وتم تدريج مفردات الاختبار وفقاً لصعوبتها.

واستخدم هدسون (Hudson, 1989) مداخل متنوعة للقياس في تقدير مستوى القدرة في اللغة الإنجليزية. حيث استخدم اختبار مكون من صورتين

متكافئتين وطبق الاختبار على ١٣٠ طالباً وطالبة من عدة دول وتوصل إلى أن نظرية الاستجابة للمفردة كانت أكثر دقة من الطرائق التقليدية في تحليل مفردات الاختبار.

وأجرت شادية عبدالعزيز (١٩٩٦) دراسة لاستخدام نموذج راش في بناء بنك للأسئلة في مقرر علم النفس التعليمي، وبلغ حجم العينة ٢٢٣٦ طالبة، وطبقت عليهن مفردات اختبارية في علم النفس التعليمي بلغ عددها ٣٤٤ مفردة من نوع الاختيار من متعدد ذات ثلاثة بدائل. وتوصلت الدراسة إلى تدرج هذه الأسئلة (بعد حذف بعض الأفراد والمفردات) على متصل واحد يحقق استقلالية القياس.

واستخدم القرشي (Elkorashy, 1995) نموذج راش في تدرج بنود اختبار أوتيس لينون للقدرة العقلية العامة. وطبق الاختبار على ٥٩٩ طالباً وطالبة بالمدارس الثانوية والجامعية بالكويت. وتم تدرج بنود الاختبار بعد حذف ٦ بنود لتدني معاملات تمييزها وحذف ٢٤ بنداً لعدم ملائمتها لنموذج راش، وقد ارتفع معامل ثبات الاختبار للتدرج النهائي لعدد ٥٠ بنداً إلى ٠.٩٧.

وقام جرين وكليفز (Green & Kluever, 1991) بتدرج اختبار المصفوفات المتتابعة لرافن بصورتيه الملونة والعادية. وبلغ حجم العينة ٢٦٩ طالباً وطالبة، ١٥١ موهوباً بالصف السابع. وتم تدرج مفردات الاختبار باستخدام نموذج راش. وتوصلت النتائج إلى أن الصورة الملونة كانت أفضل من الصورة العادية. واستخدمت منى ربيع (٢٠٠٠) نموذج راش في تطوير اختبار المصفوفات المتتابعة لرافن. وبلغ حجم العينة ١٤١١ طالباً وطالبة تتراوح أعمارهم بين ٦-١٣ سنة من عدة محافظات مصرية. وتم حذف ١١ بنداً غير ملائم للنموذج وتم تدرج ٤٩ بنداً موزعة على المجموعات الخمس للاختبار.

وأجرى حمدي أبو جراد (٢٠٠٥) دراسة لإعداد اختبار الاستعداد العلمي للتنبؤ بالنجاح الأكاديمي بالجامعة. وأعد عدة مقاييس للمهارات المعرفية والإتجاهات العلمية وعمليات العلم الأساسية والتكاملية. وطبق الأدوات على

٢٩٤ طالباً وطالبة بالثانوية العامة بغزة. تم تدريج مقاييس الاستعداد العلمي باستخدام نموذج راش، وتوصلت الدراسة إلى ان مقاييس الإستعداد العلمي كانت من أفضل المنبئات للنجاح في كلية الصيدلة.

كما أجرى عبدالناصر أحمد (٢٠٠٧) دراسة لبناء مقاييس للقبول بالجامعات المصرية. وبلغت عينة الدراسة ٥٥٣ طالباً وطالبة طبق عليهم عدة مقاييس للإستعداد اللغوي والرياضي والإستعداد للغة الإنجليزية. وقد تم تدريج الاختبارات باستخدام نموذج راش، حيث تم حذف ٣٦ فرداً وثلاث عبارات عند تدريج مقياس الإستعداد اللغوي. وحذف ٣٠ فرداً و ٦ عبارات من مقاييس الإستعداد الرياضي، وحذف ٤٦ فرداً و ٨ عبارات من مقياس استعداد اللغة الإنجليزية. وتوصلت الدراسة إلى إرتفاع معاملات الثبات من ٠,٦٩ ، ٠,٦٣ ، ٠,٦٣ ، إلى ٠,٩٧ ، ٠,٩٧ ، ٠,٩٧ .

إعداد اختبار اللغة العربية لطلبة الدراسات العليا

تُعد الكفاية اللغوية عاملاً مهماً للنجاح في الدراسات العليا، فالقدرة على فهم المعلومات المقروءة والاستنتاج منها، وتجميع المعلومات والأفكار من مصادر عدة ووضعها في محتوى جديد، يُعد أمراً ضرورياً لطلبة الدراسات العليا. فهم يستخدمون ذلك كثيراً في بحوثهم الفصلية وفي كتابة التقارير وكذلك في إجراء الأطروحات العلمية.

وبالرجوع إلى الدراسات السابقة واختبارات SAT ، GRE ، TOEFL ودراسة مكوناتها، تم تحديد المقصود من اختبار اللغة العربية لطلبة الدراسات العليا (تخصصات العلوم الإنسانية) بجامعة الكويت، وما يجب أن يقيسه كاختبار للقبول، وما يجب أن يتضمنه مثل هذا الاختبار. وعليه تم تحديد الهدف من الاختبار، والمهارات اللغوية اللازمة للقياس ويجب أن تتوافر في طلبة الدراسات العليا وذلك على غرار ما هو موجود في الاختبارات المشار إليها وخاصة اختبار التوفل TOEFL.

وتم تحديد الهدف من الاختبار بأن يكون طالب الدراسات العليا قادراً على: فهم النصوص التي يدرسها، والاستنتاج منها، والتعبير عنها بأسلوب آخر مراعيًا سلامة الصياغة وقواعد اللغة العربية. ويكون ذلك من خلال اختبار يحتوي موضوعات متنوعة تُغطي عدداً من تخصصات العلوم الإنسانية. ويتم التركيز في الاختبار على ثلاثة محاور هي: القدرة على الفهم، والقدرة على استخدام قواعد اللغة العربية، والقدرة على التفكير والتعبير عن الأفكار.

المنهج وإجراءات الدراسة

اختبار اللغة العربية لقبول طلاب الدراسات العليا

تم تحديد ثلاثة محاور أساسية يجب أن يشملها اختبار اللغة العربية لقبول طلبة الدراسات العليا بكلّيات العلوم الإنسانية بجامعة الكويت وهي: القدرة على الفهم وتتضمن استنباط معاني الكلمات وفهم المقروء وتلخيصه، والقدرة على استخدام قواعد اللغة العربية، والقدرة على التفكير وتتضمن صياغة الأفكار والاستنتاج والتحليل والربط بينها وإعادة صياغتها.

من ثم اختيار موضوعات متنوعة تُغطي مختلف تخصصات العلوم الإنسانية وصياغة بعض من فقراتها، ووضع أسئلة على هذه الموضوعات من نوع الاختيار من متعدد على أن تقيس القدرات المشار إليها آنفاً، كما تم الالتزام بأن يكون لكل سؤال من أسئلة الاختيار من متعدد خمسة بدائل وذلك بهدف تقليل أثر التخمين.

وتمت مناقشة الموضوعات المختارة والأسئلة ومراجعتها في ضوء أسس القياس النفسي من حيث الصياغة ودقة التعبير ومدى قياسها للجوانب المحددة المطلوب قياسها.

وانتهى إلى إعداد ١٣٠ سؤالاً تغطي المكونات الخمسة للاختبار وبيانها

كما يلي:

جدول رقم (١)

تصنيف بنود اختبار اللغة العربية وفقاً للمكونات الخمسة

عدد البنود	المحور
٤٤	١- فهم المقروء
٥١	٢- استخدام قواعد اللغة
٣٥	٣- التفكير والتعبير عن الأفكار
١٣٠	المجموع

تجريب الاختبار

أولاً - التجريب الاستطلاعي:

يتطلب إعداد الاختبارات الجيدة تجريبها على عينة مناسبة من المجتمع المستهدف لاستخدام الاختبار ثم إجراء التعديلات اللازمة وإعادة التجريب مرة أخرى ليخرج الاختبار في أفضل صورة له حتي يقيس ما هو محدد لقياسه. ولذلك فإن أهداف التجريب الاستطلاعي هي:

- ١ - تعرف مدى ملاءمة موضوعات الاختبار وأسئلته لطلبة الدراسات العليا من كليات العلوم الإنسانية.
 - ٢ - تعرف مدى وضوح التعليمات وكتابة البيانات الأولية للمستجيبين. وحساب الزمن المناسب لتطبيق الاختبار.
 - ٣ - حساب الخصائص السيكومترية للاختبار، وهي معاملات السهولة والتمييز، إضافة إلى معاملات الصدق والثبات.
- وقد تم وضع جميع الأسئلة المشار إليها بالجدول رقم (١) وعددها ١٣٠ سؤالاً في اختبار واحد بهدف تجريبه على عينة من طلبة الدراسات العليا.

عينة التجريب الاستطلاعي:

تم تجريب الاختبار (١٣٠ بنداً) على عينة حجمها ٨٦ (١٠ طلاب، ٧٦ طالبة) من طلبة الدراسات العليا ببرامج ماجستير التربية الدرجة الموحدة

وبرنامج علم النفس بكلية العلوم الاجتماعية. حيث تم تطبيق الاختبار بمعرفة الباحثين وبالاتفاق مع بعض أساتذة المقررات بالبرامج المذكورة.

وقد تم تحليل البيانات باستخدام برامج SPSS لدرجات الاختبار (١٣٠ بنداً)، وبلغ المتوسط الحسابي (٦٠,١٧) والانحراف المعياري (١٤,٧٦) ومعامل الثبات (٠,٨٩). كما أن متوسط زمن تطبيق الاختبار بلغ ٧٢,٢٤ دقيقة بمتوسط قدره ٠,٥٦ دقيقة للسؤال الواحد.

وبدراسة معاملات الصعوبة والتمييز وُجد أن بعض الأسئلة مرتفعة الصعوبة وبعضها منخفضة التمييز، وتراوحت معاملات السهولة بين ٠,٢٤ - ٠,٩٤ ومعاملات التمييز بين ٠,٣٣ - ٠,٥٦ ولذلك تم حذف الأسئلة منخفضة السهولة وضعيفة التمييز وكان عددهم ٣٠ بنداً. وتم إجراء تحليل البيانات مرة أخرى لعدد ١٠٠ بنداً، ونتج عن التحليل أن معاملات السهولة تراوحت بين ٠,٠٧ - ٠,٩٤ ومعاملات التمييز بين ٠,١٠ - ٠,٥٩ وأدى ذلك إلى ارتفاع معامل كرونباك ألفا إلى ٠,٩١. وتم تحديد ١٦ بنداً من ١٠٠ بند لتعديل صياغتها بهدف تحسين معاملات الصعوبة والتمييز. ويوضح جدول رقم (٢) عدد البنود التي تم حذفها والبنود المعدلة.

جدول رقم (٢)

عدد البنود التي تم تجريبيها والبنود المحذوفة والمعدلة

العناصر	عدد البنود	عدد البنود المحذوفة	عدد البنود المعدلة	مجموع البنود بعد التعديل
١- فهم المقروء	٤٤	٩	٦	٣٥
٢- استخدام قواعد اللغة	٥١	١٧	٣	٣٤
٣- التفكير والتعبير عن الأفكار	٣٥	٤	٧	٣١
المجموع	١٣٠	٣٠	١٦	١٠٠

وتم حساب العلاقة بين الدرجة الكلية للاختبار مع المعدل التراكمي (لطلبة الدراسات العليا) وبلغ معامل الارتباط ٠,٤٦ على التوالي، وهو معامل ارتباط دال ويشير إلى صدق تلازمي للاختبار، كما أن هذه العلاقة تعني أن

درجات اللغة العربية يُمكنها أن تتبأ بالمعدل التراكمي لطلبة الدراسات العليا. وهذه النتائج الأولية جيدة وتدل على أن بنود الاختبار بعد تجريبيها وتثقيحها وتعديل صياغة بعضها تكون صالحة للإستخدام مع طلبة الدراسات العليا.

تعديل الاختبار

بعد تجريب الاختبار المكون من ١٣٠ بنداً تم حذف ٣٠ بند وتعديل ١٦ بنداً من بنود الاختبار بناء على نتائج التجريب الاستطلاعي، ومن ثم أصبح الاختبار مكوناً من ١٠٠ بنداً في صورته النهائية المستخدمة لعملية التقنين وحساب المعايير على طلبة الدراسات العليا من كليات العلوم الإنسانية.

ثانياً - تقنين الاختبار

١ - إجراءات تقنين الاختبار

تم اعداد الاختبار في صورته النهائية لإجراء عملية التقنين وهو مكون من ١٠٠ بند من نوع الاختيار من متعدد ولكل بند خمسة بدائل، وروعي أن تتوزع الاجابات الصحيحة عشوائياً على أن يتساوي ظهور كل بديل كإجابة صحيحة وذلك لتقليل التخمين، وتم تحديد زمن التطبيق بـ ٦٠ دقيقة للإجابة عن أسئلة الاختبار بناء على متوسط زمن التجريب الاستطلاعي، ويتم حساب زمن التطبيق بعد الانتهاء من شرح التعليمات وتدوين البيانات الأولية. واستخدمت ورقة للإجابة يُمكن قراءتها بالماسح الضوئي وتصحيحها آلياً وإدخال بياناتها في برنامج SPSS تمهيداً لتحليل البيانات واعداد جداول المعايير.

وتمت مخاطبة كلية الدراسات العليا بشأن تقنين الاختبار والسماح بالتطبيق على طلبة الدراسات العليا في كليات العلوم الإنسانية والاتفاق على آلية التطبيق بتحديد يوماً معيناً لكل كلية لتطبيق الاختبار بمعرفة الباحثين. وقد تم التطبيق خلال الفترة من ٢٠٠٨/٣/٩ وحتى ٢٠٠٨/٥/٥.

٢ - عينة التقنين

تم تطبيق الاختبار المكون من ١٠٠ بند على عينة حجمها ٣٣٥ طالبا

وطالبة - من مجتمع الدراسة وهو ٤١٠ طلاب وطالبة - (١٠٨ طلاب، ٢٢٧ طالبة) ببرامج الماجستير والدبلوم من تخصصات العلوم الإنسانية (التربية ٧٩، الشريعة ١١٣، الآداب ١٠٥، العلوم الاجتماعية ٢٦، العلوم الإدارية ١٢)، وبلغ عدد المتفرغين للدراسة ١٥٦ (٣٥ طالباً، ١٢١ طالبة) وغير المتفرغين ١٧٩ (٧٣ طالباً، ١٠٦ طالبة). ويوضح الجدولان المرفوقان (٣، ٤) توصيف عينة التقنين.

جدول رقم (٣)

توصيف العينة وفقاً للكلية والنوع

الكلية / النوع	ذكور	إناث	المجموع
التربية	٨	٧١	٣٩
الشريعة	٥١	٦٢	١١٣
الآداب	٣٤	٧١	١٠٥
العلوم الاجتماعية	٥	٢١	٢٦
العلوم الإدارية	١٠	٢	١٢
المجموع	١٠٨	٢٢٧	٣٣٥

جدول رقم (٤)

توصيف العينة وفقاً للنوع والدوام

الدوام / النوع	ذكور	إناث	المجموع
كلي	٣٥	١٢١	١٥٦
جزئي	٧٣	١٠٦	١٧٩
المجموع	١٠٨	٢٢٧	٣٣٥

ثالثاً - النتائج

تم تحليل الإجابات على بنود الاختبار باستخدام الطريقة التقليدية لحساب معاملات الصعوبة والتمييز لكل بند، وحساب معامل الثبات بطريقة كرونباك ألفا لعدد ١٠٠ بند، وبلغ ٠,٨٤، وقد تراوحت معاملات الصعوبة بين ٠,٠٦ - ٠,٩٨، ومعاملات التمييز بين ٠,٠٠١ - ٠,٤٣، ونستنتج من ذلك أنه على الرغم من ارتفاع معامل الثبات إلا أنه توجد بنود مرتفعة السهولة

والصعوبة وبنود ضعيفة التمييز، وعليه تم حذف ١٧ بنداً من الاختبار وأصبح مكوناً من ٨٣ بنداً، وتراوحت معاملات السهولة بين ٠,١٥ - ٠,٩١ ومعاملات التمييز بين ٠,١١ - ٠,٤٥ وبلغ معامل كرونباك ألفا ٠,٨٦ لعدد ٨٣ بنداً.

صدق الاختبار

تم حساب معامل الارتباط بين درجات الاختبار المعدل مع درجات المعدل التراكمي وبلغ ٠,٢٣ وهو دال عند ٠,٠٠١ لكنه منخفض. وقد يرجع ذلك إلى انخفاض تباين درجات المعدل التراكمي لطلاب الدراسات العليا، ويدل هذا الارتباط الدال على صدق الاختبار، كما يحدد إمكانية تنبؤ درجات الاختبار بالمعدل التراكمي لطلبة الدراسات العليا في حين كان معامل ارتباط درجات TOEFL مع المعدل التراكمي غير دالة وقريبة من الصفر.

وتم حساب درجات المحاور الثلاثة لاختبار اللغة العربية المعدل، ومن ثم حساب معاملات الارتباط بينها وبين الدرجة الكلية (وهو ما يسمى بالاتساق الداخلي) ويتضح من جدول رقم (٥) أن معاملات الارتباط مرتفعة ودالة عند ٠,٠٠١ مما يدل على صدق الاختبار.

جدول رقم (٥)

معاملات الاتساق الداخلي لدرجات محاور اختبار اللغة العربية

المتغير	فهم المقروء	استخدام قواعد اللغة	التفكير والتعبير عن الأفكار
اختبار اللغة العربية (المعدل)	٠,٨٨	٠,٨٤	٠,٧٩
١ - فهم المقروء		٠,٦١	٠,٥٧
٢ - استخدام قواعد اللغة			٠,٤٦
٣ - التفكير والتعبير عن الأفكار			

❖ جميع الارتباطات دالة عند ٠,٠٠١.

وتم إجراء تحليل لكل من درجات الاختبار والمعدل التراكمي لمعرفة الفروق وفقاً للنوع والادوام والكلية وتوضح الجداول أرقام (٦، ٧، ٨) هذه الفروق.

جدول رقم (٦)

الفروق وفقاً للنوع في درجات الاختبار والمعدل التراكمي

المتغير	ذكور		إناث		قيمة ت	مستوى الدلالة
	متوسط	ع	متوسط	ع		
اختبار اللغة العربية (المعدل)	٥١,٠٦	١٠,٣٧	٤٧,٨٩	١٠,٥٧	٢,٥٩	٠,٠١
١ - فهم المقروء	١٨,٣٩	٤,٣٧	١٧,٠٥	٤,٤٥	٢,٥٩	٠,٠١
٢ - استخدام قوائد اللغة	١٦,٦٦	٤,٣٨	١٥,٢٤	٠,٣٠	٢,٧٤	٠,٠٠٦
٣ - التفكير والتعبير عن الأفكار	١٦,٠٢	٣,٩٣	١٥,٦٠	٠,٢٤	٠,٩٦	غ دال
المعدل التراكمي	٣,٥١	٠,٣٩	٣,٦٠	٠,٣٣	٢,٠٣-	٠,٠٤٣

يوضح جدول رقم (٦) وجود فروق دالة بين متوسطي الجنسين في درجات جميع المتغيرات حيث يتضح أن متوسطات الذكور أعلى من الإناث في درجات اختبار اللغة العربية المعدل وأبعاده، في حين كان متوسط الإناث أعلى من الذكور في المعدل التراكمي.

جدول رقم (٧)

الفروق وفقاً للدوام في درجات الاختبار والمعدل التراكمي

المتغير	دوام كلي		دوام جزئي		قيمة ت	مستوى الدلالة
	متوسط	ع	متوسط	ع		
اختبار اللغة العربية المعدل	٥٠,٢٣	١٠,١٤	٤٧,٧٦	١٠,٨٧	٢,١٤	٠,٠٣٣
١ - فهم المقروء	١٧,٧٧	٤,٢٤	١٧,٢٣	٤,٦٤	١,١١	غ دال
٢ - استخدام قواعد اللغة	١٦,٣٧	٤,٤٤	١٥,١١	٤,٤٢	٢,٥٨	٠,٠١
٣ - التفكير والتعبير عن الأفكار	١٦,١٠	٣,٤٦	١٥,٤٢	٣,٩٤	١,٦٨	غ دال
المعدل التراكمي	٣,٦٣	٠,٢٩	٣,٥١	٠,٣٩	٢,٩٦	٠,٠٠٣

ويتبين من جدول رقم (٧) وجود فروق دالة بين متوسطي الدوامين الكلي والجزئي في درجات اختبار اللغة العربية المعدل والمحور الثاني والمعدل التراكمي لصالح مجموعة الدوام الكلي.

كما تم إجراء تحليل تباين أحادي لمعرفة الفروق بين متوسطات الكليات في درجات اختبارات اللغة العربية والمعدل التراكمي.

جدول رقم (٨)

قيم ف ودلالاتها الإحصائية من تحليل التباين بين الكليات
في درجات الاختبار والمعدل التراكمي

المعدل التراكمي	المحور الثالث	المحور الثاني	المحور الأول	الاختبار المعدل	الكلية/ المتغير
٣,٦٥	١٥,٤٤	١٣,٦٦	١٦,٤٨	٤٥,٥٨	التربية
٣,٦٢	١٥,٩٣	١٧,٦١	١٩,١٩	٥٢,٧٣	الشرعية
٣,٤٩	١٦,٠٢	١٥,٩٩	١٧,٣٨	٤٩,٣٩	الآداب
٣,٦٩	١٤,٧٧	١٣,٤٥	١٤,١٥	٤٢,٢٧	العلوم الاجتماعية
٢,٨٥	١٥,٤٢	١٣,٥٨	١٦,٠٨	٤٥,٠٨	العلوم الإدارية
❖١٤,٤٤	٠,٨١	❖١٣,٦٧	❖١٠,٠٢	❖٩,٤٩	قيمة ف

(❖) دال عند ٠,٠٠١

يتضح من جدول رقم (٨) وجود فروق دالة بين متوسطات الكليات في درجات الاختبارات والمعدل التراكمي عند مستوى دلالة مرتفع (٠,٠٠١)، ولمعرفة اتجاهات الفروق تم إجراء المقارنات المتعددة بين المتوسطات باستخدام اختبار دنكان Duncan ووُجد ما يلي:

- ١ - متوسطات درجات اختبار اللغة العربية المعدل: متوسط الشرعية أعلى من متوسطات بقية الكليات عدا الآداب، ومتوسط درجات الآداب أعلى من العلوم الاجتماعية، ويُعد هذا دليلاً على صدق الاختبار لأنه من المألوف أن طلاب الشرعية أكثر استخداماً للغة العربية مقارنة بالكليات الأخرى.
- ٢ - متوسطات المعدل التراكمي: متوسط درجات طلبة علوم الإدارة هي الأقل ولا توجد فروق بين متوسطات درجات الكليات الأخرى.

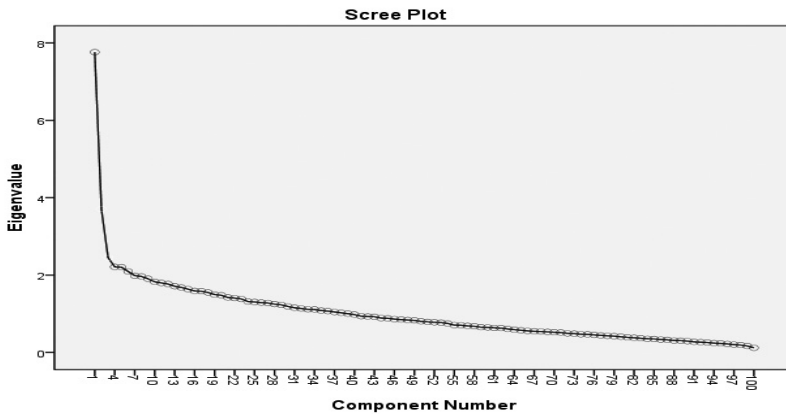
وبإجراء تحليل تباين أحادي بين البرامج وُجد أن متوسطات درجات طلبة برامج الترجمة واللغة العربية والشرعية هي أعلى المتوسطات بمستوى دلالة

٠,٠٠١ وأقل المتوسطات لبرنامجي العلوم السياسية والإرشاد النفسي. وهذا دليل آخر على صدق الاختبار.

نتائج التحليل باستخدام نموذج راش

يتكون اختبار اللغة العربية من (١٠٠) سؤال تم تدرجهم باستخدام نموذج راش Rasch Model وذلك على عينة من طلاب الدراسات العليا حجمها ٣٣٥ طالباً وطالبة. وفيما يلي استعراض للبيانات الإحصائية التي أسفر عنها تحليل استجابات الطلاب على أسئلة الاختبار وذلك باستخدام بعض البرامج الإحصائية المناسبة وتشمل: SPSS و BILOG-MG و WINSTEPS.

أولاً - التحقق من أحادية البعد: نظراً لأن أحادية البعد (وجود عامل واحد رئيسي) أحد الشروط الأساسية لنموذج راش وتُعني أن أسئلة الاختبار تكون متجانسة فيما بينها وتقيس في أساسها نفس الصفة، فقد تم حساب مصفوفة الارتباط الرباعي بين أسئلة الاختبار وإخضاعها للتحليل العاملي الاستكشاف في طريقة المكونات الأساسية لهوتلنج والتدوير المتعامد للمحاور بطريقة الفاريماكس. وبالرغم من تشبع الأسئلة على عدة عوامل إلى أن الشكل رقم (١) يوضح العلاقة بين كل سؤال وقيمة الجذر الكامن المقابل scree plot وتدل على وجود عامل واحد رئيسي.



شكل رقم (١): العلاقة بين كل سؤال وقيمة الجذر الكامن

ثانياً - التحقق من توازي المنحنيات المميزة للبند: للتحقق من توازي المنحنيات المميزة للبند تم حساب معامل الارتباط الثنائي point-biserial correlation لأسئلة الاختبار كما هو موضح في الجدول رقم (٩). وبالرغم من اختلاف قيمة الارتباط إلا أننا يمكن القول بأنها متقاربة.

جدول رقم (٩)

معاملات تمييز أسئلة الاختبار باستخدام معامل الارتباط الثنائي

السؤال	معامل الارتباط الثنائي	السؤال	معامل الارتباط الثنائي	السؤال	معامل الارتباط الثنائي	السؤال	معامل الارتباط الثنائي	السؤال	معامل الارتباط الثنائي
q_1	.226	q_21	.212	q_41	.314	q_61	.289	q_81	.270
q_2	.347	q_22	.405	q_42	.204	q_62	.260	q_82	.242
q_3	.213	q_23	.256	q_43	.326	q_63	.273	q_83	.251
q_4	.328	q_24	.249	q_44	.231	q_64	.196	q_84	.191
q_5	.259	q_25	.266	q_45	.216	q_65	.283	q_85	.206
q_6	.245	q_26	.198	q_46	.231	q_66	.215	q_86	.232
q_7	.340	q_27	.232	q_47	.189	q_67	.265	q_87	.310
q_8	.261	q_28	.260	q_48	.216	q_68	.306	q_88	.268
q_9	.298	q_29	.350	q_49	.228	q_69	.267	q_89	.260
q_10	.235	q_30	.188	q_50	.206	q_70	.287	q_90	.222
q_11	.281	q_31	.241	q_51	.358	q_71	.299	q_91	.251
q_12	.234	q_32	.211	q_52	.291	q_72	.211	q_92	.267
q_13	.273	q_33	.235	q_53	.222	q_73	.334	q_93	.283
q_14	.230	q_34	.184	q_54	.185	q_74	.301	q_94	.320
q_15	.364	q_35	.254	q_55	.266	q_75	.396	q_95	.375
q_16	.299	q_36	.188	q_56	.193	q_76	.247	q_96	.434
q_17	.198	q_37	.360	q_57	.196	q_77	.248	q_97	.431
q_18	.218	q_38	.231	q_58	.238	q_78	.374	q_98	.276
q_19	.211	q_39	.169	q_59	.238	q_79	.224	q_99	.227
q_20	.319	q_40	.391	q_60	.241	q_80	.249	q_100	.184

ثالثاً - معاملات صعوبة الأسئلة بعد تدريسها باستخدام نموذج راش

جدول رقم (١٠)

معاملات الصعوبة لأسئلة الاختبار باللوجيت والخطأ المعياري لها

الأسئلة	الصعوبة باللوجيت	الخطأ المعياري (ع)	الأسئلة	الصعوبة باللوجيت	الخطأ المعياري (ع)	الأسئلة	الصعوبة باللوجيت	الخطأ المعياري (ع)
1	-2.470	0.161	36	-0.369	0.100	71	-0.124	0.100
2	0.893	0.114	37	0.103	0.101	72	0.144	0.101
3	0.393	0.104	38	0.534	0.106	73	1.083	0.120
4	-0.252	0.099	39	1.697	0.145	74	0.134	0.101
5	2.566	0.205	40	-3.247	0.222	75	0.024	0.100
6	2.445	0.194	41	-0.203	0.099	76	-0.998	0.106
7	0.457	0.105	42	-0.942	0.105	77	0.382	0.104
8	-0.124	0.100	43	-1.973	0.135	78	-0.803	0.103
9	-0.232	0.099	44	1.740	0.147	79	-0.026	0.100
10	-0.428	0.100	45	-0.065	0.100	80	-0.679	0.102
11	-1.705	0.124	46	-1.784	0.127	81	1.718	0.146
12	-1.042	0.107	47	-4.014	0.315	82	-0.498	0.100
13	3.100	0.260	48	-1.254	0.111	83	2.003	0.162
14	-2.394	0.156	49	-1.884	0.131	84	0.546	0.106
15	-1.170	0.109	50	0.134	0.101	85	1.718	0.146
16	-0.588	0.101	51	-0.261	0.099	86	1.464	0.134
17	-1.230	0.110	52	-0.409	0.100	87	-1.042	0.107
18	-0.164	0.099	53	-1.181	0.109	88	0.298	0.103
19	-0.998	0.106	54	0.034	0.100	89	-0.488	0.100
20	1.429	0.132	55	0.073	0.100	90	1.429	0.132
21	0.999	0.117	56	-1.850	0.130	91	-1.329	0.113
22	0.073	0.100	57	3.877	0.375	92	-0.986	0.106
23	1.901	0.156	58	-1.784	0.127	93	-1.705	0.124
24	-0.478	0.100	59	-0.301	0.099	94	-0.115	0.100

تابع/ جدول رقم (١٠)

معاملات الصعوبة لأسئلة الاختبار باللوجيت والخطأ المعياري لها

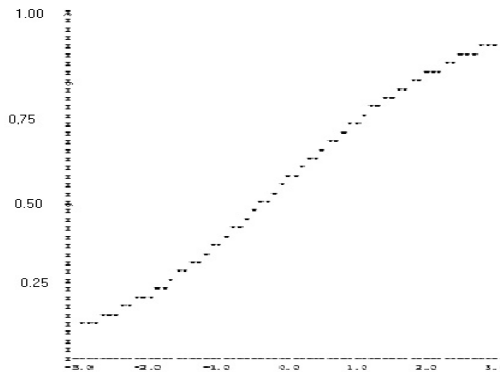
الأسئلة	الصعوبة باللوجيت	الخطأ المعياري (ع)	الأسئلة	الصعوبة باللوجيت	الخطأ المعياري (ع)	الأسئلة	الصعوبة باللوجيت	الخطأ المعياري (ع)
25	0.602	0.107	60	5.820	0.973	95	0.906	0.114
26	-0.537	0.100	61	3.171	0.268	96	1.500	0.136
27	2.445	0.194	62	-0.845	0.104	97	-1.571	0.120
28	-1.720	0.125	63	1.377	0.130	98	-1.816	0.128
29	-0.986	0.106	64	-0.213	0.099	99	1.538	0.137
30	2.336	0.186	65	-0.588	0.101	100	-0.409	0.100
31	-1.420	0.115	66	1.877	0.155	متوسط الصعوبة	صفر	
32	-1.487	0.117	67	-0.468	0.100	ع	1,29	
33	0.014	0.100	68	0.371	0.104			
34	-0.751	0.102	69	1.111	0.121			
35	-0.359	0.100	70	0.298	0.103			

ويتضح من جدول رقم (١٠) تباين مستويات صعوبة أسئلة الاختبار حيث تراوحت من الأسئلة السهلة جداً (مثل مفردة رقم ٤٧) إلى أسئلة صعبة (مثل مفردة رقم ٦٠). كما أن قيم الخطأ المعياري منخفضة في معظم البنود. كما يُلاحظ أن معظم أسئلة الاختبار داخل نطاق الصعوبة من -٢ إلى +٢ لوجيت مع وجود بعض الأسئلة السهلة القليلة التي تقل معامل صعوبتها عن -٢ وعدد قليل من الأسئلة التي تزيد صعوبتها عن +٢ أو تقل عن -٢. ووجود معظم الأسئلة داخل نطاق الصعوبة (-٢، +٢) طبيعي حيث أن معظم قدرات الطلاب تكون داخل هذا النطاق.

ويتضح أيضاً من الجدول رقم (١٠) أن معاملات صعوبة الأسئلة تغطي مدى واسع من الصعوبة عند الأخذ في الاعتبار الخطأ المعياري لها، كما يتضح أن الخطأ المعياري للأسئلة الأصعب أكبر من الخطأ المعياري لمعاملات الصعوبة للأسئلة الأسهل وهو أمر منطقي؛ حيث إنه كلما زادت مستويات الصعوبة قل عدد المستجيبين بشكل صحيح على المفردة، وبالتالي يزيد الخطأ المعياري للمفردة.

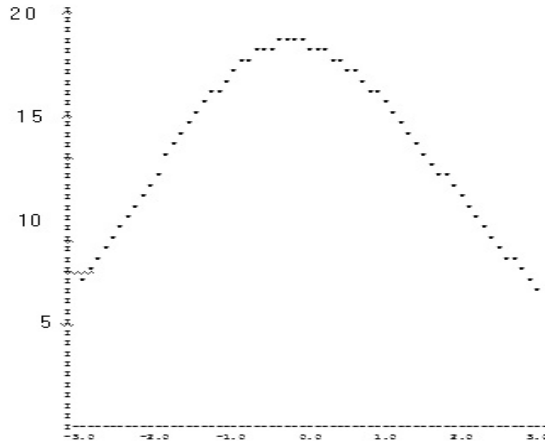
واستخدم برنامج WINSTEPS في تحليل بيانات الاختبار ثلاث مرات، الأولى لحساب معاملات صعوبة البنود ومستويات تقدير القدرة لكل بند وكذلك معامل ثبات بنود الاختبار ومعامل ثبات تقدير قدرات الأفراد على البنود. ويتضح من التحليل أن متوسط قدرة الطلاب منخفض (٠,٣٥) لوجيت بانحراف معياري صغير نسبياً (٠,٥٩)، مما يدل على تجانس عينة الطلاب من حيث القدرة وتمركزهم حول قدرات ضعيفة نسبياً. كما كان متوسط صعوبة الأسئلة (صفر) لوجيت بانحراف معياري كبير نسبياً (١,٢٩) بالمقارنة بنظيره الخاص بصعوبة الأسئلة. مما يدل على تغطية صعوبة الأسئلة لمدي واسع من الصعوبة ولكنه لا يقابل المدى الذي تغطيه قدرات الطلاب.

ويتضح من النتائج أيضاً تقارب توزيع قدرات الأفراد عن توزيع صعوبات الأسئلة. فكما أوضحنا مسبقاً بأن متوسط صعوبة أسئلة الاختبار كان مساوياً الصفر بمتوسط قدرة الأفراد ٠,٣٥ لوجيت على تدرج الاختبار، مما يؤدي لارتفاع الانحراف المعياري لتقدير كل من صعوبة الأسئلة وقدرات الأفراد. كما يتضح من جدول رقم (١٠) الأسئلة مرتفعة السهولة والتي تتفق وقدرات الأفراد حيث كانت معظم أسئلة الاختبار داخل نطاق الصعوبة (١-، ١+) وبالتالي كان بعض الأسئلة بعيدة نسبياً عن مستوى قدرات الطلاب.



شكل رقم (٢) : المنحنى المميز للاختبار

يوضح شكل رقم المنحني المميز للاختبار وهو يأخذ الشكل القياسي للمنحنيات المميزة، وأهم ما يميز هذا المنحني هو انخفاض مستويات التخمين نسبياً على الاختبار حيث لم تزيد عن الحد المتوقع (٠,٢٥) عند الطلاب (أقل من ٢- لوجيت) (على افتراض أن أسئلة الاختبار هي من نوع الاختيار من متعدد منخفضي القدرة ذو البدائل الأربع)، كما ارتفعت احتمالات الإجابة الصحيحة على الاختبار مع زيادة قدرة الطلاب وفقاً للدالة اللوغاريتمية المتوقعة للعلاقة بين قدرة الأفراد واحتمالات اجاباتهم الصحيحة على الأسئلة.



شكل رقم (٣): منحني معلومات الاختبار Test Information Curve

ويوضح شكل رقم (٣) منحني معلومات الاختبار ويأخذ الشكل الاعتدالي المتوقع لمثل هذه المنحنيات، كما كانت أكبر قيمة للمعلومات عن مستوى قدرات الأفراد وفقاً للاختبار بدقة مرتفعة داخل نطاق القدرة (-١، +١) وهو ما كان متوقعاً نظراً لأن معظم أسئلة الاختبار غطت هذا النطاق. وانخفضت دقة المعلومات التي يمكن الحصول عليها من الاختبار بشكل كبير مع زيادة مستوى القدرة عن +١ أو انخفاض مستوى القدرة عن -١، وهو ما يؤكد ضرورة إضافة أسئلة تغطي تلك المستويات من القدرة.

ويتم في التحليل الثاني باستخدام برنامج WINSTEPS استبعاد الأفراد الحاصلين على الدرجة النهائية أو صفر في إجاباتهم على الاختبار، كما يتم استبعاد الأفراد غير الملائمين لنموذج راش Rasch باستخدام محكي الملاءمة Infit و Outfit عند مستوى ± 2 ، وتعني $+2$ أن نمط إجابة الفرد يختلف عما هو متوقع منه بسبب التسرع أو الإهمال أو التخمين، أما من يتجاوز الحد -2 فيعني أن استجابته غير واقعية وقد ترجع إلى البطء أو الحرص الشديد أو الغش. وبلغ عدد الأفراد غير الملائمين الذين تم حذفهم ٢٦ فرداً من مجموع ٣٣٥ حجم العينة بنسبة تصل إلى ٧,٧٦٪ وهي تُعد نسبة مرتفعة إلى حد ما لكنها مقبولة، وتراوحت معاملات الملاءمة Infit للأفراد بين ٣,٧ إلى (-٢,١) بينما تراوحت معاملات عدم الملائمة Outfit بين ٤,٧ و (-٢,٩).

أما التحليل الثالث فيتم فيه حذف البنود غير الملائمة باستخدام محك الملاءمة ٢ وقد أدى ذلك إلى حذف ١٩ بنداً من بنود الاختبار المئة مما يؤدي إلى تجانس بنود الاختبار في قياس القدرة في اللغة العربية. ويعني تجاوز بنود الاختبار محك الملاءمة $+2$ بوجود عيب في الصياغة أو أن البنود تعرف متغيراً آخر يختلف عن باقي بنود الاختبار، أما تجاوز محك الملاءمة -2 فإنها تعني أن البنود تقيس متغيراً آخر مرتبطاً بالمتغير موضع القياس، ولذا تم استبعاد هذه البنود غير الملائمة لأسس القياس الموضوعي. ويوضح جدول رقم (١١) ملخصاً لنتائج التحليلات الثلاثة المذكورة.

جدول رقم (١١)

ملخص نتائج تحليل بيانات الاختبار باستخدام برنامج WINSTEPS

التحليل الأول	التحليل الثاني	التحليل الثالث	
٣٣٥	٣٠٩	٣٠٩	عدد الأفراد
١٠٠	١٠٠	٨١	عدد البنود
صفر	صفر	صفر	متوسط الصعوبة
١,٢٩	١,٣٨	١,٥٢	متوسط الانحراف المعياري للصعوبة

تابع/ جدول رقم (١١)

ملخص نتائج تحليل بيانات الاختبار باستخدام برنامج WINSTEPS

التحليل الأول	التحليل الثاني	التحليل الثالث	
٠,٣٥	٠,٤٠	٠,٥٠	متوسط القدرة
٠,٥٩	٠,٥٩	٠,٦٤	متوسط الانحراف المعياري للقدرة
٠,٨٣	٠,٨٣	٠,٨١	معامل ثبات البنود
٠,٩٩	٠,٩٩	٠,٩٩	معامل ثبات تقدير القدرة

يتضح من جدول رقم (١١) أن متوسط تقدير صعوبة بنود الاختبار هو صفر لوجيت في التحليلات الثلاث ويتراوح متوسط الانحراف المعياري لتقديرات الصعوبة بين ١,٢٩ و ١,٥٢ وبلغ معامل ثبات الاختبار ٠,٨٣ في التحليلين الأول والثاني ومعامل ثبات تقدير قدرات البنود ٠,٩٩ كما تراوحت متوسطات تقدير القدرة بين ٠,٣٥ و ٠,٥٠ أما معامل ثبات الاختبار بعد حذف ١٩ بنداً فقد بلغ ٠,٨١.

وعملية استبعاد البنود غير الملائمة والإبقاء على الملائم منها يؤكد صدق بنود الاختبار في قياس القدرة المراد قياسها. كما أن حذف الأفراد غير الملائمين والبنود غير الملائمة يحقق صدق بنود الاختبار وصدق تدرج قدرات الأفراد (كاظم، ١٩٨٨) و(أبو جراد، ٢٠٠٥) و(أحمد، ٢٠٠٧).

كما تم حساب الفروق بين مستوى صعوبة كل بندين متتالين في التدرج ومقارنته بمجموع الخطأ المعياري لهما؛ حيث إنه إذا كان الفرق بين الصعوبتين أقل من مجموع الخطأ المعياري فهذا يعني عدم وجود فراغ على متصل القدرة المقاسة، أما إذا كان الفرق بين الصعوبتين أكبر من مجموع الخطأ المعياري لهما فهذا يعني وجود فراغ بين البندين على متصل القدرة المقاسة مما يتطلب إضافة بنود جديدة لقياس القدرة. جدول رقم (١٢) يوضح ترتيب البنود والفروق بين معاملات صعوبة كل بندين متتالين، ومجموع الخطأ المعياري لهما والقرار المتخذ بشأنهما.

جدول رقم (١٢)

الفرق بين مستوى صعوبة كل بندين متتاليين ومقارنته

بمجموع الخطأ المعياري لهما

ترتيب البنود	الفرق بين معاملات صعوبة كل بندين	مج الخطأ المعياري لهما	القرار	ترتيب البنود	الفرق بين معاملات صعوبة كل بندين	مج الخطأ المعياري لهما	القرار
٣٢				٩٠	٠.٠٥	٠.٢٥	...
٧٠	٠.٤٩	٠.٥٨	...	١٤	٠.٠١	٠.٢٦	...
٧٧	٠.٥٨	٠.٤٦	فراغ	٨١	٠.٠٥	٠.٢٦	...
٨٤	٠.١٢	٠.٤٠	...	٤١	٠.٠٢	٠.٢٦	...
٥٤	٠.١٨	٠.٣٨	فراغ	٨٦	٠.٠١	٠.٢٦	...
٤٦	٠.٥٨	٠.٣٣	...	٣	٠.٠٨	٠.٢٦	...
٤٤	٠.٠٤	٠.٣٠	...	٤٨	٠.٠٢	٠.٢٦	...
٣٨	٠.٠٧	٠.٣٠	...	١	٠.٠٢	٠.٢٦	...
٣٣	٠.٢١	٠.٢٩	...	١٨	٠.٠١	٠.٢٦	...
٧٩	٠.١١	٠.٢٨	...	٢٩	صفر	٠.٢٦	...
٦٢	٠.٠٢	٠.٢٨	...	٤٠	٠.٠٧	٠.٢٦	...
٣٥	٠.٠٢	٠.٢٨	...	٨٢	صفر	٠.٢٦	...
١١	٠.٠٧	٠.٢٧	...	٣٠	٠.٠٧	٠.٢٦	...
٣١	٠.٠٤	٠.٢٦	...	٥٥	٠.٠٢	٠.٢٦	...
١٩	٠.٢٠	٠.٢٦	...	٤	٠.٠٤	٠.٢٦	...
٨٥	٠.١٣	٠.٢٦	...	٦١	صفر	٠.٢٦	...
٢٦	٠.١٠	٠.٢٥	...	٧٤	صفر	٠.٢٦	...
٩	٠.٠٨	٠.٢٤	...	٧٨	٠.٠٣	٠.٢٧	...
٨٨	٠.٠١	٠.٢٤	...	١٢	٠.٠٤	٠.٢٨	...
٨	٠.٠٢	٠.٢٤	...	٥٠	صفر	٠.٢٨	...
٢٧	٠.٠٣	٠.٢٤	...	٨٩	٠.٠٦	٠.٢٨	...
١٠	٠.٠١	٠.٢٤	...	٦٠	٠.٠٢	٠.٢٨	...
٤٩	٠.٠٢	٠.٢٤	...	١٦	٠.١٠	٠.٢٨	...
٩٢	٠.٠٩	٠.٢٤	...	٤٣	٠.٢٣	٠.٢٩	...
٦٤	٠.١٣	٠.٢٤	...	٢٠	٠.٠٣	٠.٣٠	...
٩٨	صفر	٠.٢٤	...	٢٢	صفر	٠.٣٠	...
١٠٠	٠.١٧	٠.٢٤	...	٨٧	٠.٠٧	٠.٣١	...
٣٤	٠.١٥	٠.٢٤	...	٢	٠.٠٥	٠.٣٢	...

تابع/ جدول رقم (١٢)
الفروق بين مستوى صعوبة كل بندين متتاليين ومقارنته
بمجموع الخطأ المعياري لهما

ترتيب البنود	الفروق بين معاملات صعوبة كل بندين	مج الخطأ المعياري لهما	القرار	ترتيب البنود	الفروق بين معاملات صعوبة كل بندين	مج الخطأ المعياري لهما	القرار
٩١	صفر	٠,٢٤	...	٧٦	٠,١٩	٠,٣٣	...
٩٤	٠,٠٥	٠,٢٤	...	٥	٠,٠٥	٠,٣٤	...
٣٩	٠,١٤	٠,٢٤	...	٢٣	٠,١٣	٠,٣٥	...
٤٧	٠,٠٣	٠,٢٤	...	٧١	٠,١٠	٠,٣٧	...
٥٣	صفر	٠,٢٤	...	٦٦	٠,١٣	٠,٣٨	...
٥٦	٠,١٣	٠,٢٤	...	٢٤	٠,٠٥	٠,٣٩	...
٦٨	٠,٠٣	٠,٢٤	...	٢٥	٠,٠٤	٠,٤٠	...
٩٣	صفر	٠,٢٤	...	٥٩	٠,٢٣	٠,٤٢	...
٧٣	٠,٠١	٠,٢٤	...	٣٦	٠,٥٧	٠,٥١	فراغ
٨٠	٠,٠٢	٠,٢٤	...	٤٢	٠,١٨	٠,٦٠	...
١٣	٠,١٠	٠,٢٤	...	٦٧	صفر	٠,٦٢	...
٦	٠,٠٢	٠,٢٤	...	٦٥	١,٧٥	١,٠٢	فراغ
٢٨	صفر	٠,٢٤	...				

ويتضح من جدول رقم (١٢) أن الفروق بين الصعوبات أقل من مجموع الخطأ المعياري ما عدا في حالتين في بداية متصل القدرة بين البندين ٧٠ و ٧٧ وبين البندين ٥٤ و ٤٦ وحالتين في نهاية القدرة بين البندين ٥٩ و ٣٦ وبين البندين ٦٧ و ٦٥ ويدل هذا على جودة تدريج بنود الاختبار، ومن الممكن إضافة بنود إلى بداية التدرج ونهايته لتغطية الفراغات الموضحة، ويكون ذلك من خلال إعداد بنك لأسئلة اختبار اللغة العربية بما يساعد في توفير صور متكافئة للاختبار عند الاستخدام.

جدول رقم (١٣)

معايير الاختبار تبين الدرجة الخام والقدرة باللوجيت والدرجة المعدلة للقدرة

الدرجة الخام	القدرة باللوجيت	الدرجة المعدلة	الدرجة الجامعية	الدرجة الخام	القدرة باللوجيت	الدرجة المعدلة	الدرجة الجامعية
١٥	٢,٠١-	٨٢	١١٠	٤٣	٠,١٥	١٠١	٤٤٦
١٦	١,٩٠-	٨٣	١٢٦	٤٤	٠,٢٢	١٠٢	٤٥٦
١٧	١,٨٠-	٨٤	١٤٢	٤٥	٠,٢٩	١٠٣	٤٦٧
١٨	١,٧١-	٨٥	١٥٧	٤٦	٠,٣٦	١٠٣	٤٧٨
١٩	١,٦١-	٨٦	١٧١	٤٧	٠,٤٣	١٠٤	٤٨٩
٢٠	١,٥٣-	٨٦	١٨٥	٤٨	٠,٥٠	١٠٥	٥٠٠
٢١	١,٤٤-	٨٧	١٩٨	٤٩	٠,٥٧	١٠٥	٥١١
٢٢	١,٣٦-	٨٨	٢١١	٥٠	٠,٦٤	١٠٦	٥٢٢
٢٣	١,٢٧-	٨٩	٢٢٤	٥١	٠,٧١	١٠٦	٥٣٤
٢٤	١,١٩-	٨٩	٢٣٦	٥٢	٠,٧٩	١٠٧	٥٤٥
٢٥	١,١٢-	٩٠	٢٤٩	٥٣	٠,٨٦	١٠٨	٥٥٧
٢٦	١,٠٤-	٩١	٢٦٠	٥٤	٠,٩٤	١٠٨	٥٦٩
٢٧	٠,٩٦-	٩١	٢٧٢	٥٥	١,٠٢	١٠٩	٥٨١
٢٨	٠,٨٩-	٩٢	٢٨٤	٥٦	١,١٠	١١٠	٥٩٣
٢٩	٠,٨٢-	٩٣	٢٩٥	٥٧	١,١٨	١١١	٦٠٦
٣٠	٠,٧٥-	٩٣	٣٠٦	٥٨	١,٢٦	١١١	٦١٩
٣١	٠,٦٧-	٩٤	٣١٧	٥٩	١,٣٥	١١٢	٦٣٢
٣٢	٠,٦٠-	٩٥	٣٢٨	٦٠	١,٤٤	١١٣	٦٤٦
٣٣	٠,٥٣-	٩٥	٣٣٩	٦١	١,٥٣	١١٤	٦٦٠
٣٤	٠,٤٧-	٩٦	٣٥٠	٦٢	١,٦٢	١١٤	٦٧٥
٣٥	٠,٤٠-	٩٦	٣٦١	٦٣	١,٧٢	١١٥	٦٩٠
٣٦	٠,٣٣-	٩٧	٣٧١	٦٤	١,٨٢	١١٦	٧٠٥
٣٧	٠,٢٦-	٩٨	٣٨٢	٦٥	١,٩٢	١١٧	٧٢١
٣٨	٠,١٩-	٩٨	٣٩٢	٦٦	٢,٠٣	١١٨	٧٣٨

تابع/ جدول رقم (١٣)

معايير الاختبار تبين الدرجة الخام والقدرة باللوجيت والدرجة المعدلة للقدرة

الدرجة الخام	القدرة باللوجيت	الدرجة المعدلة	الدرجة الجامعية	الدرجة الخام	القدرة باللوجيت	الدرجة المعدلة	الدرجة الجامعية
٣٩	-٠,١٢	٩٩	٤٠٣	٦٧	٢,١٤	١١٩	٧٥٦
٤٠	-٠,٠٦	٩٩	٤١٤	٦٨	٢,٢٦	١٢٠	٧٧٤
٤١	٠,٠١	١٠٠	٤٢٤	٦٩	٢,٣٩	١٢١	٧٩٤
٤٢	٠,٠٨	١٠١	٤٣٥	٧٠	٢,٥٢	١٢٣	٨١٥

وتم تحويل الدرجات الخام إلى درجات باللوجيت ودرجة معدلة (متوسطها=١٠٠ بانحراف معياري=٩) وكذلك درجة جامعية (متوسطها=٥٠٠ بانحراف معياري=١٠٠) وهي مشابهة لدرجات اختبار التوفل. وتراوحت الدرجات الخام على الاختبار (المكون من ٨١ بنداً) بين ١٥-٧٠ وتراوح مقابلها بالدرجة المعدلة بين ٨٢-١٢٣، أما الدرجات الجامعية تراوحت بين ١١٠-٨١٥. وحيث إن متوسط قدرات الأفراد باللوجيت ٠,٥٠ فإن الدرجة الجامعية المقابلة لها هي ٥٠٠، وبذلك نقترح أن تكون درجة القبول لطلبة الدراسات العليا على الاختبار هي درجة خام ٤٨ التي تساوي الدرجة الجامعية ٥٠٠.

ويوضح جدول رقم (١٣) معايير الاختبار حيث يتضمن الدرجة الخام والقدرة باللوجيت والدرجة المعدلة (الجامعية) للقدرة.

استنتاج عام

نستنتج مما سبق أن اختبار اللغة العربية (المعدل) الذي يتكون من ٨١ بنداً يتمتع بدرجة ثبات عالية بين ٠,٨١-٠,٨٦ ومستوى صدق جيد يتضح من معاملات الاتساق بين درجات أبعاده المرتفعة والدالة. كما أن معاملات ارتباط درجات اختبار اللغة العربية (المعدل) مع المعدل التراكمي دالة عند مستوى ٠,٠٠١ وهذا يعني أن درجات اختبار اللغة العربية يمكنها التنبؤ بالمعدل التراكمي، كما يتضح صدق وثبات الاختبار من جودة تدرج بنوده وقياسها

لمستويات القدرة وفقاً لنموذج راش؛ بما يشير إلى مستوى ثقة متميز في إمكانية استخدام درجات اختبار اللغة العربية المعدل كاختبار قبول لطلبة الدراسات العليا (إلى جانب درجات TOEFL) بدرجة قطع تُعادل ٤٨ نصف خطأ معياري، حيث تتراوح بين ٤٦ - ٥٠ لجميع الكليات وهي تعادل درجات معدلة من ١٠٣ - ١٠٦ كما تعادل درجات جامعية من ٤٧٨ - ٥٢٢.

التوصيات البحثية المستقبلية

- هناك بعض الدراسات التي يجب القيام بها للتأكد من صلاحية معايير ودرجات القطع لاختبار اللغة العربية لقبول الطلبة ببرامج الدراسات العليا بكليات العلوم الانسانية بجامعة الكويت منها على سبيل المثال وليس الحصر:
- ١ - إعداد صور مختلفة من الاختبار من خلال إنشاء وتطوير بنك للأسئلة Item Bank ومن ثم التحقق من صدق تكافؤ الصور Test Equivalency.
 - ٢ - إجراء دراسة تنبؤية على عينة أكبر تغطي كافة التخصصات الأدبية وحساب معادلة انحدار لكل تخصص على حدة.
 - ٣ - إجراء دراسة استكشافية باستخدام اختبار اللغة العربية على عينة من طلبة الدراسات العليا بالكليات العملية.

Using Rasch Model in Standardizing Arabic Test for Admitting Students to Graduate Programs in Colleges of Humanities at Kuwait University

Dr. Fawzyiah A. Hadi

Prof. Salah A. Mourad

College of Education
Kuwait University

Abstract

The study aims to develop a standardized Arabic test for Graduate students of colleges of humanities at Kuwait University applying to CGS. Arabic Test of 130 items was constructed and pilot tested on 86 students from Graduate School. As a result, 30 items were deleted and 16 items were modified. The modified version of the test was administered to a sample of 335 students and data were analyzed both by using traditional and Rasch model techniques. Results of traditional technique come up to deleting 17 items, and Alpha Cronbach coefficient was .86. Internal consistency coefficients ranged between .46-.88 and all correlations between test scores and GPA were statistically significant at .001.

Rasch model was used to calibrate test items (100 items) and being unsuitable to the Rasch model, 26 individuals and 19 items were deleted. The reliability indices were .81, .99 for test items and individuals abilities respectively. Ability latitude ranged between ± 2 logit providing a good indicator. Norms and cutoff were determined by using logit, standardized score, and college score (similar to TOFEL) for students admitting in the graduate programs at Kuwait University.

المراجع

- ١ - أبو جراد، حمدي يونس (٢٠٠٥). الإستعداد العلمي والتنبؤ بالنجاح في بعض كليات الجامعة: دراسة سيكومترية في نظرية الاستجابة للمفردة. رسالة دكتوراة غير منشورة، معهد البحوث والدراسات العربية، جامعة الدول العربية.
- ٢ - أبوعلام، رجاء والديب، فتحي (١٩٨٥). القوة التنبؤية لمجموع درجات امتحان شهادة الثانوية العامة بأداء الطلبة في جامعة الكويت. المجلة التربوية، ٢٢(٤)، ٥٧-١١.
- ٣ - أحمد، عبدالناصر عبدالحليم (٢٠٠٧). دراسة سيكومترية لبناء مقياس للقبول بالجامعات المصرية باستخدام نظرية الاستجابة للمفردة وعلاقته بالنجاح الأكاديمي. رسالة دكتوراه غير منشورة، كلية التربية: جامعة الفيوم.
- ٤ - السيد، فوقية عبدالفتاح (١٩٩٥). التنبؤ بنجاح طلاب كليات الهندسة في ضوء استعداداتهم الأساسية وقدراتهم الطائفية نحو دراسة الهندسة. رسالة دكتوراة غير منشورة، معهد الدراسات والبحوث التربوية، جامعة القاهرة: جمهورية مصر العربية.
- ٥ - الشافعي، محمد منصور (١٩٩٦). أثر طرق معادلة درجات الاختبار وضوابط اختيار العينة على تدريج بنوك الأسئلة باستخدام نموذج راش. رسالة دكتوراه غير منشورة، كلية التربية، جامعة المنصورة.
- ٦ - شريف، نادية (١٩٨٨). دراسة مقارنة لمستويات النجاح والقيمة التنبؤية في المرحلة الثانوية والمرحلة الجامعية لطلبة نظام المقررات والنظام التقليدي. المجلة التربوية، ١٥(٤)، ١٦١-١٩١.

- ٧ - الصرifi، عبدالله عبدالغني (١٩٨٨). التنبؤ بالتحصيل الدراسي لطلاب جامعة أم القرى من درجاتهم في الثانوية العامة وبعض المتغيرات الأخرى. مجلة البحوث النفسية والتربوية، السنة الخامسة عشر، العدد الثاني.
- ٨ - الطنطاوي، منى ربيع (٢٠٠٠). دراسة سيكومترية حول تطوير اختبار المصفوفات المتتابعة لرافن باستخدام نموذج راش. رسالة ماجستير غير منشورة، كلية البنات، جامعة عين شمس.
- ٩ - عبدالعزيز، شادية (١٩٩٦). استخدام نموذج راش في بناء بنك للأسئلة لمقرر في علم النفس التعليمي وتحديد الدرجات الفاصلة المقابلة للتقديرات الجامعية. رسالة ماجستير غير منشورة، كلية البنات، جامعة عين شمس: جمهورية مصر العربية.
- ١٠ - علام، صلاح الدين محمود (١٩٨٦). تطورات معاصرة في القياس النفسي والتربوي. الكويت: جامعة الكويت.
- ١١ - كاظم، أمينة محمد (١٩٨٨). استخدام نموذج راش في بناء اختبار تحصيلي في مقرر علم النفس وتحقيق التفسير الموضوعي للنتائج. في: أنور الشرقاوي وآخرون، اتجاهات معاصرة في القياس والتقويم النفسي والتربوي. القاهرة: الأنجلو المصرية.
- ١٢ - مراد، صلاح وسليمان، أمين (٢٠٠٥). الاختبارات والمقاييس في العلوم النفسية والتربوية. (طبعة ٢). الكويت: دار الكتاب الحديث.
- 13 - Bastruk, R. (1999). The relationship of Graduate Record Examination Aptitude Test scores and Graduate School Performance of international students at the United States universities. Paper presented a the Annual Conference of the Mid-Western Educational Research Association. Chicago: Illinios, Oct. 13-16.
- 14 - Bridgeman, B., McCamley-Jenkins, L.& Ervin, N. (2000). Prediction of Freshman grade point average from the Revised and Recentered SAT1: Reasoning Test. College Board Research Report (ED446593).

- 15 - Cooper, T. C. (1987). Foreign Language Study and SAT-Verbal scores. **Modern Language Journal**, 71: 381-387.
- 16 - Cooper, T.C., Yanosky II, D.J., Wisenbaker, J.M., Jahner, D., Webb, E. & Wilbur, M. L. (2008). Foreign Language Learning and SAT Verbal Scores Revisited. **The American Council on the Teaching of Foreign Languages**, 41(2), 200-217.
- 17 - Edy, P.A. (1981). The effect of Foreign Language Study in high school on verbal ability as measured by the Scholastic Aptitude test-verbal. Center for Applied Linguistics. Washington: D.C.
- 18 - El-Korashy, A. (1995). Validity and reliability of an Arabic modified version of the Otis-Lennon mental ability test. **Psychological Reports**, 55(5), 753-763.
- 19 - Forbes, D. W. (1986). The Rasch model as a practical and effective procedure for educational measurement. Paper presented at the Annual Meeting of NCME, San Francisco.
- 20 - Geiser, S. (2008a). Back to the basics: In defense of achievement (and achievement tests) in college admissions. CHSE, Center for Studies in Higher Education, University of California, Berkeley.
< <http://chse.berkeley.edu/> > .
- 21 - Geiser, S. (2008b). Not So Fast: A second opinion on a University of California Proposal to Endorse the new SAT. CHSE, Center for Studies in Higher Education, University of California, Berkeley.
< <http://chse.berkeley.edu/> > .
- 22 - Green, K. E. & Kluever, R. C. (1991). Component identification and item difficulty of Raven's items. Paper presented at the Annual Meeting of NCME, Chigaco.
- 23 - Henning, G. (1984). Advantages of Latent Trait measurement in language testing. **Language Testing**, 1, 123-133.
- 24 - Hudson, T. D. (1989). Measurement approaches in the development of functional ability language tests: Norm-referenced, criterion-referenced and item response theory decisions. Ph.D thesis, University of California, Los Angeles,
- 25 - Kirkup, C., Schagen, I., Wheeler, R., Morrison, J. & Whetton,

- C.(2007). Use of an Aptitude Test in university entrance- A validity study, Relationships between SAT scores, attainment measures and background variables. National Foundation for Educational Research (NFER). Research Report No. 846.
- 26 - Menson, R. P.; Patelis, T. & Doyle, A. (2009). New England's States of College Readiness. The College Board.
- 27 - Morrison, T. & Morrison, M. (1995). A meta-analytic assessment of the predictive validity of the quantitative and verbal components of the Graduate Record Examination with Grade Point Average representing the criterion of graduate success. **Educational and Psychological Measurement**, 55, 309-16.
- 28 - Powers, D.E., Swinton, S. S. & Carlson, A. B. (1977). A factor analytic study of the GRE Aptitude Test, GRE Board. Professional Report GREB No. 75.
- 29 - Ray, S. A. (2008). Standardized Tests as measurement of achievement: Does the high school assessment program (HASP) measure up? MA thesis, Southern Wesleyan University.
- 30 - Tanner, D. E. (2003). Admissions and Placement Testing: Enough is enough. Research Report 143. ERIC, ED 478175.
- 31 - Wright, B., Mead, R. & Bell, S. (1980). BICAL: Calibrating items with Rash model, Research Memorandum. Dept. of Education, University of Chicago: Illinois.
- 32 - Wright, B. & Stone, M.(1979). **Best test design: a handbook for Rash measurement**. Chicago: Meta Press.

