استخدام نظرية الاستجابة على الفقرة في تحليل الاستجابات على اختبار المهارات الكمية لدى طالب الجامعة

د. حمزة محمد دودين

قسم علم النفس - كلية العلوم الإنسانية والاجتماعية

جامعة الإمارات العربية المتحدة

الملخص

تشمل المهارات الكمية مهارات الرياضيات بفروعها كالهندسة، والجبر، والحساب، والإحصاء، بالإضافة إلى القدرة على قراءة الجداول البيانية والرسومات والأشكال المتصلة بالتعابير الكمية والعددية وفهمها وتحليلها. وتعتبر المهارات الكمية إحدى مكونات برنامج التعليم الجامعي العام، وهي لا شك ضرورية لطالب الجامعة اليوم بغض النظر عن تخصصه الأكاديمي. وكما يجب أن تحرص الجامعة على تنمية المهارات الكمية لدى طلبتها، فإن عليها كذلك أن تحرص على تقويمها بشكل موضوعي ودقيق ومستمر. وتعرض هذه الدراسة للمنهجية العلمية التي استخدمتها جامعة الإمارات العربية المتحدة في إعداد وتطوير اختبار لقياس المهارات الكمية لطلبتها. وقد قامت هذه المنهجية على الاعتماد على المخرجات التعليمية في إعداد وبناء فقرات الاختبار، كما اعتمدت نظرية الاستجابة على مستوى الفقرة (Item Response Theory (IRT)) كإطار لتحليل استجابات الطلبة. وقد تكون الاختبار من ٤٠ فقرة من نوع الاختيار من متعدد في خمسة مجالات كمية هي: الحساب، والجبر، والهندسة، والأشكال البيانية والجداول، والإحصاء والاحتمالات. وقد جرى تطبيق الاختبار على عينة من ١٠٠٥ طلاب في جامعة الإمارات العربية المتحدة. وحيث إن مثل هذا الاختبار سيستخدم بشكل دوري مما يعني استهلاك فقراته بسرعة، فقد تم اعتماد آلية خاصة في نظرية الاستجابة على مستوى الفقرة تسمى تصميم الارتكاز (Anchor Design) لإضافة فقرات جديدة بشكل مستمر، بحيث يمكن بناء نماذج جديدة

ومتكافئة من الاختبار كلما لزم الأمر. ويمكن للمنهجية المعروضة في هذه الدراسة أن تطبق لقياس قدرات، أو استعدادات، أو مهارات أخرى كثيرة لدى طلبة الجامعة كالقدرة اللغوية ومهارات التفكير وغيرها، بالإضافة إلى سهولة تطبيقها في الاختبارات التحصيلية.

مقدمة:

يشكل برنامج التعليم الجامعي العام المكون الأساسي المشترك في التعليم الجامعي. ويتكون هذا البرنامج من مجموعة من المساقات والخبرات التعليمية التي تُوفَّر لكل طلبة الجامعة. وتهدف برامج التعليم الجامعي العام إلى بناء وتنمية مجموعة مشتركة من المهارات والمعارف والاتجاهات والقيم الأساسية المنسجمة مع أهداف الجامعة وخطتها الاستراتيجية. ومن المهارات التي تحرص كثير من الجامعات على تنميتها في جميع طلابها مهارات الاتصال الشفوي أو الكتابي، ومهارات استخدام التكنولوجيا، ومهارات التفكير، والمهارات الكمية. تشمل المهارات الكمية مهارات الرياضيات بفروعها كالهندسة، والجبر، والحساب، والإحصاء، بالإضافة إلى القدرة على قراءة الجداول البيانية والرسومات والأشكال المتصلة بالتعابير الكمية والعددية وفهمها وتحليلها.

وتعتبر المهارات الكمية ضرورية لطالب الجامعة اليوم، بغض النظر عن تخصصه الأكاديمي، وذلك لمساعدته في فهم الكثير من أمور المعرفة المتصلة بالكميات وتحليلها، والتعابير الرياضية، والإحصائية مما يساعده على النجاح في دراسته الجامعية، كما أنها تساعده في تنمية مهاراته الأخرى وخصوصاً قابليته للتعلم مدى الحياة (Hambure & Jackson, 2000)، ورفع مستواه المهني، وإعداده للحياة العملية. إن الكثير من المهارات الأساسية ترتبط بطريقة أو بأخرى بالمعارف والمهارات الكمية؛ ومنها على سبيل المثال: مهارات حل المشكلات، والبحث العلمي، والمبادأة، والتفكير بأنواعه الناقد والإبداعي والمجرد، والاستخدام الفعال للتكنولوجيا، بالإضافة إلى التواصل مع الآخرين وفهمهم. كذلك تساعد مهارات الطالب الكمية وتفكيره الرياضي في التغلب على

المشاكل والصعوبات العملية والعلمية بأسلوب علمي فعال؛ إذ إنها تنمي عند الطالب القدرة على الفهم والتحليل، وتحديد عناصر المشكلة وعلاقاتها، ثم ربط ذلك ببدائل الحلول الممكنة (Hockings, 2001).

لقد أصبحت المهارات الكمية في الوقت الحاضر مكوناً أساسياً من مكونات برنامج التعليم الجامعي العام وهدفاً مهماً من أهداف مؤسسات التعليم العالي (Bauman & Martin, 1995). وكما يجب أن تحرص الجامعة على تنمية المهارات الكمية لدى طلبتها، فإن عليها كذلك أن تحرص على تقويمها بشكل موضوعي ودقيق ومستمر. وهذا ما يستدعي بناء وتطوير أدوات قياس مناسبة كاختبار المهارات الكمية. وإضافة إلى ما سبق، فإن مثل هذا الاختبار سيستخدم بشكل دوري مما يعني استهلاك فقراته بسرعة، وبالتالي الحاجة إلى إعداد اختبار جديد في كل مرة يطبق فيها. من أجل ذلك كان لا بد من اعتماد آلية محددة لإضافة فقرات جديدة بشكل مستمر، بحيث يمكن بناء نماذج جديدة ومتكافئة من الاختبار كلما لزم الأمر.

وتعرض هذه الدراسة للمنهجية العلمية التي استخدمت في جامعة الإمارات العربية المتحدة في إعداد وتطوير اختبار لقياس المهارات الكمية لطلبتها، وقد قامت هذه المنهجية على استخدام المخرجات التعليمية كأساس لبناء وتطوير فقرات الاختبار، كذلك تعرض هذه الدراسة للآلية التي تم من خلالها تحليل فقرات الاختبار، وهي آلية اعتمدت على نظرية الاستجابة على مستوى الفقرة (Item Response Theory (IRT)). وقد قامت بإعداد الاختبار والإشراف على تطويره لجنة من أعضاء هيئة التدريس شُكِّلت على مستوى الجامعة، وضمت في عضويتها مختصين في مجال المهارات الكمية وبناء الاختبارات. وبالإضافة إلى مشاركته في بناء الاختبار وتطويره، قام الباحث بالتحليل الإحصائي للاختبار، وبتحليل استجابات الطلبة، وفقرات الاختبار في إطار نظرية الاستجابة على الفقرة.

الإطار النظري:

بما أن بناء المهارات الكمية لدى طلبة الجامعة يتم من خلال مجموعة من المساقات والبرامج الفرعية، وعلى مدى سنوات الدراسة الجامعية، فإن من الصعب تحديد محتوى موحد يمكن الاعتماد عليه في بناء مقياس لتقويم المهارات الكمية، لذا تظهر الحاجة إلى الاعتماد على الأهداف والمخرجات التعليمية في بناء مثل هذا الاختبار أكثر من المحتوى. كذلك فقد اعتمدت هذه الدراسة على نظرية الاستجابة على مستوى الفقرة كإطار لتحليل الاختبار؛ إذ بالرغم من شيوع استخدام النظرية الكلاسيكية في القياس (Classical Test Theory)، وخصوصاً في تحليل نتائج الاختبارات وتفسيرها، إلا أنها تعاني من نقاط ضعف عدة (Stocking, 1999). ومن أهم تلك النقاط أن خصائص الاختبار وفقراته تحدد بالنسبة لمجموعة معينة من الممتحنين، كما أن خصائص الممتحنين أنفسهم تحدد أيضاً بالنسبة إلى اختبار معين وليس بشكل مطلق. ومن جهة أخرى، تفترض النظرية الكلاسيكية في القياس أن خطأ القياس المعياري (Standard Error of Measurement) متساوٍ لكل الممتحنين وهذا غير دقيق. وأخيراً وليس آخراً، فإن هذه النظرية لا تبين مدى تحصيل الطالب على مستوى الفقرة وإنما على الاختبار ككل (Hambleton, Swaminalton, & Rogers, 1991)، مما يعني نقصاً في المعلومات المتوافرة وصعوبة في تحديد مدى اكتساب الطالب لمخرجات محددة.

أما نظرية الاستجابة على الفقرة فتقوم على أساس أن هناك سمة كامنة (Latent Trait) تحدد قدرة الطالب على الاستجابة الصحيحة على فقرة ما، وأن قدرة الطالب هي العامل الوحيد الذي يحدد احتمالية هذه الاستجابة الصحيحة. كما تقوم هذه النظرية بتحديد قدرة الطالب (Ability) على نفس المقياس الذي تحدد عليه صعوبة الفقرة (Difficulty)، وبالتالي يمكن مقارنة صعوبة الفقرة بقدرة الطالب، وهذا ما لا يتم في إطار النظرية الكلاسيكية في القياس التي تربط صعوبة الفقرة بإجابة كل الممتحنين عليها كمجموعة واحدة.

أما بالنسبة للفقرة نفسها فإن هناك أكثر من خاصية لتحديد تلك الاحتمالية وذلك حسب النموذج المستخدم في معايرة الاختبار، وهناك ثلاثة نماذج مشهورة في نظرية الاستجابة على الفقرة لتمثيل الاختبارات الموضوعية وهي: النموذج الأحادي المعلم (One-Parameter Model)، ويحدد هذا النموذج خاصية واحدة للفقرة وهي الصعوبة، والنموذج الثنائي المعلم (Two-Parameter Model)، والذي يحدد خاصتين لكل فقرة هما الصعوبة والتمييز، وأما النموذج الثالث فيسمى النموذج الثلاثي المعلم (Three-Parameter Model)، وهو يحدد لكل فقرة ثلاث خصائص رئيسية هي: الصعوبة، والتمييز، والتخمين.

المخرجات التعليمية التي يقيسها الاختبار:

يهدف اختبار المهارات الكمية إلى قياس قدرة طالب الجامعة في المجال الكمي، وإلى تحديد المواضيع أو المجالات الرياضية أو الكمية التي تحتاج إلى عناية واهتمام أكبر من قبل الجامعة. ولقياس ذلك قامت لجنة الاختبار بتحديد خمسة مجالات ليشملها اختبار المهارات الكمية، وهي: الحساب، والجبر، والهندسة، والأشكال البيانية والجداول، والإحصاء والاحتمالات، كذلك تم تحديد المخرجات التعليمية التي لا بد لطالب الجامعة أن يمتلكها، مهما كان تخصصه الأكاديمي، والتي يتم تحقيقها من خلال ما توفره الجامعة للطالب في برنامج التعليم الجامعي العام من مساقات وخبرات تعليمية. والمخرجات التعليمية التي يقيسها هذا الاختبار هي:

- أن يحلّ الطالب مسائل عددية،
- أن يميّز الطالب بين الأعداد والمقادير الرياضية.
- أن يميّز الطالب بين الاشكال الهندسية ويحدد المسافات والزوايا والأبعاد فيها.
- أن يطبّق الطالب بعض المفاهيم والقوانين الرياضية في مواقف مختلفة.
- أن يطبّق الطالب بعض مفاهيم وقوانين الاحتمالات.

- أن يحلّ الطالب معادلات جبرية ويميّز العلاقات بين أجزائها،.
- أن يطبّق الطالب المفاهيم والعلاقات الرياضية في حل مسائل عملية.
- أن يستخرج الطالب بدقة معلومات محددة من جداول وأشكال بيانية ذات محتويات كمية.

إعداد فقرات الاختبار:

قامت لجنة الاختبار بإعداد فقرات محددة لقياس كل مجال من المجالات المذكورة، وبعد ذلك تم اختيار أفضل ثماني فقرات في كل مجال لتشكل في مجموعها ٤٠ فقرة، وتم الاتفاق على أن تكون فقرات الاختبار من نوع الاختيار من متعدد (٤ بدائل لكل سؤال)، ولما كان من الضروري أن تكون بدائل كل فقرة من فقرات الاختبار معدة جيداً وتمثل الاختيار المتوقع للطالب عند إجابته على السؤال، فقد تم وضع بدائل الأسئلة بطريقتين:

- إما أن يكون البديل نتاج تفكير منطقي غير صحيح يقوم به الطالب عند إجابته على السؤال المعطى، خصوصاً وأن الإجابة على سؤال كمي تتم عادة من خلال عمليات تفكير وحسابات معينة في ضوء فهم الطالب للسؤال وقدرته على الإجابة.
- وإما أن يكون البديل إحدى الإجابات الخاطئة ذات التكرار الأكبر من قبل الطلبة. وقد طُبّق الاختبار بدون بدائل على عينة عشوائية أولية مُمثّلة لطلبة الجامعة مكوّنة من ٣٦٠ طالباً وطالبة، وحُسبت تكرارات الإجابات على كل فقرة. وتم استخدام هذا التطبيق الأولي كذلك في التحقق من وضوح الفقرات، والصياغة، وسلامة اللغة، ومدى مناسبتها للطلبة وكذلك في تحديد الوقت اللازم للاختبار.

تطبيق الاختبار:

تم تطبيق الاختبار بصورته النهائية في الفصل الدراسي الثاني من العام الجامعي ٢٠٠١-٢٠٠٢ على عينة عشوائية ممثلة لطلبة جامعة الإمارات مؤلفة

من ١٠٠٥ طالب وطالبة من الكليات التالية: العلوم الإنسانية والاجتماعية، والعلوم، والتربية، والإدارة، والاقتصاد، والشريعة والقانون، ونظم الأغذية، والهندسة. أما عن نسبة توزيع الطلبة حسب الجنس فكانت ٢٩١ (٢٩٪) من الذكور و٧١٤ (٧١٪) من الإناث، وكانت نسبة تمثيل الجنسين في الاختبار قريبة إلى نسبة الجنسين في الجامعة، حيث هي حوالي ٢٤٪ ذكور مقابل ٧٦٪ إناث. (العدد الكلي لطلبة الجامعة في ذلك العام هو ٥٠٠ ، ١٥: ٧٥٠ ، ٣ ذكور مقابل ٩٣٠ ، ١١ إناث).

الخصائص السيكومترية لاختبار المهارات الكمية:

أولاً - صدق الاختبار:

١ - **الصدق الظاهري:** يقصد بالصدق الظاهري أن تبدو أداة القياس مناسبة لما صُممت لقياسه. وقد تم بناء اختبار المهارات الكمية من قبل لجنة متخصصة، ومرت عملية بنائه بمراحل مختلفة حتى تم اختيار فقرات مناسبة، ومن ثم مراجعتها وتدقيقها. وبعد أن اكتمل بناء وإعداد فقرات الاختبار تم عرضه على مجموعة من المحكمين من المختصين في تدريس المهارات الكمية كأساتذة الرياضيات والإحصاء في الجامعة، والمختصين في مجالات القياس والتقويم التربوي وبناء أدوات التقويم، وتم جمع الملاحظات التي قدمت من هؤلاء المحكمين، ومراجعة الاختبار وتقويمه في ضوئها.

٢ - **صدق المحتوى:** لا يوجد في الاختبارات المعيارية، كاختبارات القدرات مثلاً، محتوى محدد للتأكد من صدقه. إن الهدف في اختبار القدرات والذكاء هو عادة التحصيل أو الأداء المستقبلي كما يشير (Oosterhof, 1994): «إن الهدف من مثل هذه الاختبارات هو التنبؤ بالأداء المستقبلي، وبالتالي فإن الصدق يجب أن يؤكد القدرة على التنبؤ المستقبلي وليس المحتوى الذي تعلمه الطالب» (ص٥٩). وبالنسبة لاختبار المهارات الكمية، فقد تم، كما ذُكر سابقاً، تحديد المخرجات التعليمية التي يقيسها هذا

الاختبار، كذلك تم تسكين فقرات الاختبار على تلك المخرجات، أي تم ربط كل فقرة بالمخرج الذي تقيسه.

٣ - **الصدق المعياري:** إن تحديد معيار مناسب يُعد أهم وأصعب خطوة في فحص الصدق المعياري لاختبار ما (Linn and Gronlund, 2000)، وحيث إنه لم يسبق لمعظم طلاب الجامعة أن تقدموا لإجراء اختبارات قدرات كمية أو ذكاء؛ فإن المهمة في تحديد المعيار تبدو أصعب. لذا استخدم التحصيل الأكاديمي مقاساً بالمعدل التراكمي كمعيار خارجي. ولا شك أن استخدام المعدل التراكمي كمعيار لمقارنة التحصيل في اختبار كمي أمر غير دقيق تماماً، فالمعدل التراكمي قد يُمثل تحصيل الطالب في مواد ليست ذات علاقة بالتحصيل الكمي أو المهارات الرياضية، وحتى إذا وُجدت بعض المواد ذات طبيعة كمية فإن تأثيرها قد يكون قليلاً في حساب المعدل التراكمي العام.

وتم حساب معامل الارتباط بين نتائج الطلبة على اختبار المهارات الكمية والمعدل التراكمي، فكانت قيمة ارتباط بيرسون ٠,١١ فقط، ولتفسير قيمة هذا الارتباط تم حساب معامل التحديد فكان ١٪، أي أن ١٪ فقط من تحصيل الطلبة في اختبار المهارات الكمية يفسر بعلاقته الخطية بالمعدل التراكمي، وهذا يدل على ضعف العلاقة بين المقياسين. وأما عن ارتباط التحصيل في اختبار المهارات الكمية مع معدل الثانوية العامة، فقد حُسب معامل ارتباط بيرسون، وكانت قيمته ٠,٣٣. وهذا أيضاً يدل على ارتباط ضعيف. وتفسير ذلك، كما سبق وذكر، هو عدم وجود معيار خارجي مناسب لمقارنة التحصيل في اختبار المهارات الكمية.

٤ - **الصدق البنائي:** تم تقدير الصدق البنائي لهذا الاختبار من خلال إجراء التحليل العاملي باستخدام برنامج الكمبيوتر (TESTFACT)، وبينت النتائج أن هناك عاملاً واحداً رئيسياً تتجمع حوله فقرات الاختبار. واستطاع هذا العامل أن يفسّر وحده حوالي ٢٨٪ من تباين

درجات الاختبار، أما العامل الثاني فلم يفسّر أكثر من ٦٪، والثالث ٢٪. وبالتالي يمكن القول بأن الاختبار أحادي العامل. كذلك أظهرت نتائج التحليل العاملي تشبع فقرات الاختبار بمعاملات عالية على العامل الأول كما يظهر في الجدول رقم (١).

الجدول رقم (١)

معاملات تشبع فقرات الاختبار على العامل الأول

الفقرة	معامل تشبع الفقرة على العامل الأول	الفقرة	معامل تشبع الفقرة على العامل الأول	الفقرة	معامل تشبع الفقرة على العامل الأول	الفقرة	معامل تشبع الفقرة على العامل الأول
١	٠,٥٥	١١	٠,١٤	٢١	٠,٤٥	٣١	٠,٧٤
٢	٠,٦٠	١٢	٠,٤٣	٢٢	٠,٥٤	٣٢	٠,٧٥
٣	٠,٤٨	١٣	٠,٤٢	٢٣	٠,٦٩	٣٣	٠,٧٠
٤	٠,٤٣	١٤	٠,٥٨	٢٤	٠,٦٢	٣٤	٠,٤٩
٥	٠,٣٠	١٥	٠,٤٣	٢٥	٠,٦٢	٣٥	٠,٥٦
٦	٠,٤٢	١٦	٠,٥٥	٢٦	٠,٢٠	٣٦	٠,٦٨
٧	٠,٤٨	١٧	٠,٥٩	٢٧	٠,٥٤	٣٧	٠,٣٢
٨	٠,٣٣	١٨	٠,٥٩	٢٨	٠,٥٠	٣٨	٠,٧٠
٩	٠,٤٢	١٩	٠,٥١	٢٩	٠,٦٨	٣٩	٠,٦٠
١٠	٠,٤٤	٢٠	٠,٥٢	٣٠	٠,٦٦	٤٠	٠,٦٤

ثانياً - ثبات الاختبار:

وجد أن معامل الثبات الداخلي (كرونباخ ألفا) لاختبار المهارات الكمية ٠,٨٩. وعلى الرغم من عدم وجود معيار أو قيمة محددة لقبول ثبات الاختبار، إلا أن بعض أخصائي القياس يرى أن قيمة ثبات داخلي بحدود ٠,٦٠-٠,٨٠ يُعد مقبولاً في الاختبار العادي (المؤلف من حوالي ٤٠ فقرة) (Henderson, Morris, & Gibbon, 1987; Oosterhof, 1994).

تحليل نتائج الاختبار في إطار نظرية الاستجابة على الفقرة:

إن تحديد نموذج نظرية الاستجابة على الفقرة المناسبة لاختبار ما ليس بالأمر السهل. فكما هو معروف، أن من الصعب، من الناحية العملية، إيجاد اختبار يطابق أحد النماذج تماماً، حيث إن النموذج الأحادي المعلم يفترض تساوي الفقرات في قدراتها على التمييز وعدم وجود تخمين. والنموذج الثنائي المعلم يفترض عدم وجود تخمين. أما النموذج الثلاثي المعلم فهو الأشمل فيها والأكثر مناسبة لكثير من الاختبارات الموضوعية بسبب افتراضه الخصائص الثلاثة: الصعوبة والصدق والتخمين في تحديد احتمالية الاستجابة الصحيحة على فقرة ما. وقبل البدء بتطبيق هذه النظرية على نتائج تطبيق اختبار المهارات الكمية، هناك مجموعة من الافتراضات التي يجب توافرها في الاختبار ونتائجه، وفي حالة عدم توافر هذه الافتراضات، فإن النتائج من عملية التحليل لن تكون دقيقة.

التحقق من افتراضات النظرية:

من الأسباب التي تجعل نظرية الاستجابة على الفقرة صعبة التطبيق في حالات عملية كثيرة، اشتراطها وجود عدد كبير من الطلاب أو المتحنين يفوق ما يتوافر عادة في قاعة الصف، إذ يجب أن لا يقل الحد الأدنى لعدد المتحنين حسب النموذج الأول عن ٢٠٠ طالب، ويحتاج النموذج الثالث إلى ما لا يقل عن ١٠٠٠ طالب للحصول على نتائج دقيقة (Fan, 1998)، وعموماً كلما زاد عدد الطلاب، كانت النتائج أكثر ثباتاً. وقد تم تطبيق اختبار المهارات الكمية على ما يزيد على ١٠٠٠ طالب مما يعني توافر العدد المناسب لأي نموذج من نماذج النظرية الثلاثة. بالإضافة إلى ذلك، هناك افتراضات أخرى يجب التأكد من توافرها في الاختبار قبل البدء بإجراء التحليل، وفيما يلي عرض وتطبيق لذلك على اختبار المهارات الكمية:

١ - أحادية البناء (Unidimensionality)

تفترض نظرية الاستجابة على الفقرة أن الاختبار المراد تحليله يقيس قدرة واحد فقط. وبالرغم من أن عوامل كثيرة تؤثر على نتائج اختبار ما كقلق

الاختبار، ودافعية الطالب وغيرها، إلا أن كثيراً من الاختبارات تُبنى لقياس قدرة معينة ومنها اختبار المهارات الكمية قيد الدراسة الذي يهدف إلى قياس قدرة الطالب في المهارات الكمية «القدرة الكمية». ويتم عادة التحقق من هذا الافتراض بواسطة التحليل العاملي، وقد دلت النتائج كما ذكر سابقاً على أن العامل الأول يفسّر وحده حوالي ٢٨٪ من تباين الدرجات، وأما العامل الثاني التالي فإنه لم يفسّر سوى ٦٪، وبعد إضافة عامل ثالث فإن هذا الأخير لم يفسر أكثر من ٢٪. وهذا يعني أن العامل الأول أساسي (Dominant) وذلك باستخدام المعيار الذي قدمه (Reckase, 1979) الذي يعتبر الاختبار أحادي البناء إذا فسّر عامله الأول ٢٠٪ على الأقل من تباين الدرجات.

ولما كان اختبار المهارات الكمية قد أُعد لقياس قدرات الطلبة في خمسة مجالات كمية هي: الجبر، والهندسة، والإحصاء والاحتمالات، والأشكال البيانية، والحساب، فإن السؤال الذي يطرح نفسه هنا هو كيف ندّعي أن الاختبار أحادي البناء في حين أنه يقيس خمسة مجالات كمية؟، والحقيقة أنه لا تناقض في ذلك إذا ما أخذنا بعين الاعتبار أن هناك عاملاً مشتركاً أساسياً بين هذه المجالات الخمسة وهو قياسها لمهارات كمية، وفي هذا الموضوع يقول (Smith, 1996): «من المهم التفريق بين أحادية البناء النظرية والعملية، فمن الناحية النظرية فإن مسألة الحساب تتكون من أربعة مكونات أو بناءات مثلاً وهي: الجمع، الطرح، الضرب، والقسمة. والتحليل العاملي لاختبار الحساب يمكن أن يُظهر هذه البناءات الأربعة، ولكننا من الناحية العملية نفترض أن هذه البناءات الأربعة تقيس فعلياً موضوعاً واحداً وهو القدرة الحسابية، وفي الواقع فإن من الصعب إيجاد اختبار حساب لا يجمع بينها» (ص٢٨).

٢ - مناسبة الوقت:

تفترض نظرية الاستجابة على الفقرة أن لا يكون الاختبار المعطى اختبار سرعة، بمعنى أن يكون الوقت مناسباً وكافياً لجميع الطلبة، ويتم تحديد ذلك عملياً بمعرفة نسبة الطلبة الذين أنهوا الاختبار قبل انتهاء الوقت. وبالنسبة

لاختبار المهارات الكمية، فقد وُجد من خلال التطبيق الأولي أن معظم الطلبة (ما يزيد عن ٨٠٪) قد أنهوا الاختبار قبل انتهاء الوقت، كذلك فإن عدد الطلبة الذين لم يكملوا جميع الأسئلة كان محدوداً جداً.

ويرى بعض أخصائي القياس أن توافر العامل الأول (أحادية البناء) يعني مناسبة النتائج للتحليل في إطار نظرية الاستجابة على الفقرة، والسؤال الآن؛ أي نموذج من النماذج الثلاثة في هذه النظرية أكثر ملاءمة لهذا الاختبار؟

٣ - مناسبة النتائج لأحد النماذج النظرية:

يتم التأكد من مناسبة نتائج اختبار ما لنموذج من نماذج (IRT) بطرق مختلفة منها أولاً فحص قيم تمييز الفقرات لمعرفة ما إذا كانت هذه القيم متساوية أو على الأقل متقاربة أو غير ذلك. وقيم التمييز هذه يمكن الحصول عليها بإجراء تحليل لنتائج الاختبار حسب النظرية الكلاسيكية. وقد تم إجراء هذا التحليل وتبين من خلال النتائج أن قيم تمييز الفقرات تتفاوت بشكل كبير، فمثلاً كانت أقل قيمة لمعامل الارتباط الثنائي المتسلسل الحقيقي (Point-Biserial) للفقرة ١٠ وقدرها ٠,١٢، وكانت أعلى قيمة معامل ارتباط للفقرة ٣١ وقدره ٠,٦٢، وكانت معاملات ارتباط بقية الفقرات متفاوتة، مما يعني عدم ملاءمة النموذج الأول لهذه البيانات لأنه يفترض قيم تمييز متساوية.

أما بالنسبة للتخمين، فمن المفترض أن الطالب يلجأ إلى التخمين عندما تكون الفقرة صعبة، وقد دلت نتائج التحليل حسب النظرية الكلاسيكية أن هناك عدداً من الفقرات (٢٨٪) اعتبرت صعبة، وحيث إنه لا عقوبة على التخمين، فإن الطالب يلجأ إلى اختيار عشوائي (تخمين) للإجابة على بعض الأسئلة، وحيث إن النموذجين الأول والثاني لا يفترضان وجود تخمين، فإنهما لا يناسبان نتائج هذا الاختبار.

وأما من الناحية الإحصائية، فهناك مقاييس لمدى مناسبة كل فقرة من فقرات الاختبار للنموذج المفترض، ومنها اختبار Chi-Square في برنامج

(BILOG) (Mislevy & Bock, 1990). وقد حسبت قيمة هذا الإحصائي لكل فقرة من فقرات الاختبار، وللنماذج الثلاثة، وكانت النتائج كما يلي: عدد الفقرات التي لا تتلاءم مع النموذج الأحادي المعلم ٢٣ فقرة، عدد الفقرات التي لا تتلاءم مع النموذج الثنائي المعلم ١١ فقرة، عدد الفقرات التي لا تتلاءم مع النموذج الثلاثي المعلم ٦ فقرات. ومن هذه النتائج نلاحظ أن النموذج الثالث (الذي يستخدم صعوبة الفقرة، وتمييزها، وقيمة التخمين فيها)، هو الأكثر ملاءمة لبيانات هذا الاختبار، وبناءً عليه، فقد تم اعتماد النموذج الثلاثي المعلم لتحليل النتائج.

نتائج تحليل فقرات اختبار المهارات الكمية:

تم استخدام برنامج الكمبيوتر (BILOG) لتحليل نتائج اختبار المهارات الكمية، وتمت كذلك مقارنة النتائج باستخدام برنامج (XCALIBRE)، وفي كلتا الحالتين اعتُمد النموذج ثلاثي المعلم، وكانت النتائج متقاربة. ويعرض الجدول رقم (٢) نتائج تحليل هذا الاختبار، حيث تظهر قيم صعوبة الفقرة، تمييزها، ومقدار التخمين فيها.

الجدول رقم (٢)

نتائج تحليل اختبار المهارات الكمية في إطار نظرية الاستجابة على الفقرة

رقم الفقرة	معامل التمييز (a)	معامل الصعوبة (b)	معامل التخمين (c)	رقم الفقرة	معامل التمييز (a)	معامل الصعوبة (b)	معامل التخمين (c)
١	١,١٨	٠,٥٥	٠,٢٨	٢١	١,١٠	٠,٩٣	٠,٢٨
٢	١,٢١	١,٢٠	٠,١٨	٢٢	٠,٦١	-٠,٥٣	٠,١٠
٣	٠,٥٧	-٠,٠١	٠,١٤	٢٣	١,١٣	-٠,٩٩	٠,١٤
٤	٠,٧٤	١,٤٢	٠,١٧	٢٤	٠,٧٥	٠,١٠	٠,٠٧
٥	٠,٣٢	-٠,٥٥	٠,١٥	٢٥	١,٥١	١,٠٣	٠,١٢
٦	٠,٨٤	١,١٧	٠,٢٢	٢٦	٠,٢٥	٠,١٤	٠,١٤
٧	٠,٨٤	٠,٤٦	٠,٢٨	٢٧	٠,٨٨	٠,٩١	٠,١٧

تابع/ الجدول رقم (٢)

نتائج تحليل اختبار المهارات الكمية في إطار نظرية الاستجابة على الفقرة

رقم الفقرة	معامل التمييز (a)	معامل الصعوبة (b)	معامل التخمين (c)	رقم الفقرة	معامل التمييز (a)	معامل الصعوبة (b)	معامل التخمين (c)
٨	٠,٥٨	١,٥٤	٠,١٧	٢٨	٠,٨٨	٠,٩١	٠,١٧
٩	٠,٥٠	١,٦٦-	٠,١٢	٢٩	١,٣٣	٠,٠٣-	٠,٢٠
١٠	٠,٥٩	٢,٢٠-	٠,١٥	٣٠	١,٣١	٠,٧٣	٠,١٢
١١	١,٤١	٢,٦٢	٠,٠٧	٣١	١,٣٢	٠,٣٦	٠,٠٩
١٢	٠,٥٨	٢,٠١-	٠,١٢	٣٢	١,٥٤	٠,١٣	٠,١٣
١٣	٠,٤٩	١,٦٦-	٠,١٣	٣٣	١,٥٧	٠,٤٦	٠,١٤
١٤	٠,٧٠	٠,٥٨-	٠,١٢	٣٤	٠,٥٨	٠,٧٢	٠,٠٩
١٥	٠,٤٩	٠,٢٢	٠,١٣	٣٥	٠,٨١	٠,٢٧-	٠,١٥
١٦	٠,٦٤	٠,٧٥-	٠,١٢	٣٦	١,٢٤	٠,٦٧-	٠,١٢
١٧	٠,٧٦	٠,٧٥-	٠,٢٠	٣٧	٠,٤١	٠,٤٦	٠,١٥
١٨	٠,٩٣	٠,١٩	٠,٢١	٣٨	١,٢٠	٠,٣٢-	٠,١١
١٩	٠,٧٤	٠,١٨	٠,٢٣	٣٩	٠,٩٤	٠,٦٩-	٠,١٣
٢٠	٠,٧٠	٠,١٤	٠,١٨	٤٠	٠,٨٤	٠,١٩	٠,٠٧

تمييز الفقرة:

يكون معامل تمييز الفقرة (a) موجباً إذا كان احتمال إجابة هذه الفقرة من قبل الطلبة ذوي التحصيل العالي أكبر من الطلبة ذوي التحصيل المنخفض (Richichi, 1996). وكلما زادت قيمة تمييز الفقرة، كان ذلك أفضل ودليلاً على جودتها. ويمكن القول إنه وحسبما أشار (Choi, 1992)، فإنه إذا كانت قيمة تمييز الفقرة أكبر من ١، فهذا دليل على أن الفقرة جيدة، أما إذا كانت القيمة بين ٠,٧٠ و ١، فتمييز الفقرة مقبول. وتقل جودة الفقرة إذا قلت قيمة تمييزها

عن ٠,٧٠. وباستخدام ذلك التصنيف نجد من الجدول رقم (٢) أن فقرات اختبار المهارات الكمية قد توزعت كما يلي: ١٣ فقرة (٣٣٪) ذات تمييز ممتاز، ١٤ فقرة (٣٥٪) ذات تمييز مقبول، ١٣ فقرة (٣٣٪) ذات تمييز قليل، إلا أن الملاحظ أن معاملات التمييز في معظمها ليست متطرفة (٠,٥٠ فما فوق).

صعوبة الفقرة:

لا يوجد معيار محدد لصعوبة الفقرة (b)، فالصعوبة المرغوبة قيمة نسبية ربما تتغير من اختبار لآخر، ولكن هناك قيماً معينة تحدد فيما إذا كانت الفقرة صعبة أو سهلة في الاختبارات المعيارية عموماً، فقد أشار (Choi, 1992) أيضاً إلى أنه إذا كانت قيمة صعوبة الفقرة أكبر من ٠,٥٠، فإن الفقرة تعتبر صعبة، وإذا كانت قيمتها بين ٠,٥٠ و -٠,٥٠، فإنها تعتبر متوسطة الصعوبة، أما إذا كانت ذات قيمة أقل من -٠,٥٠، فإن هذا دليل على سهولة الفقرة. وباستخدام ذلك المعيار نلاحظ أن فقرات هذا الاختبار - كما تظهر في الجدول رقم (٢) - قد توزعت كما يلي: ١٢ فقرة (٣٠٪) صعبة، و١٦ فقرة (٤١٪) متوسطة صعبة، و١٢ فقرة (٣٠٪) سهلة. ويُلاحظ من التصنيف السابق أن معظم الفقرات كانت متوسطة الصعوبة. وبمراجعة الفقرات للتعرف على القيم المتطرفة منها، نلاحظ أن الفقرتين: ٨، ١١ ذات صعوبة متطرفة، أما الفقرات: ٩، ١٠، ١٢، ١٣ فهي سهلة إلى حد كبير.

معامل التخمين:

تراوحت قيم معامل التخمين (c) بين ٠,٠٧ (أقل قيمة) و٠,٢٨ كأعلى قيمة (الجدول رقم ٢)، وعلى العموم فإن معظم قيم معامل التخمين كانت صغيرة، حيث وجد أن معامل التخمين في ٢٦ فقرة كان أقل من ٠,١٥، وهذه القيمة أقل من القيمة النظرية المتوقعة للتخمين وهي ٠,٢٥ في حالة وجود أربعة بدائل للسؤال. وبشكل عام ومن خلال استعراض نتائج تحليل اختبار المهارات حسب نظرية الاستجابة على الفقرة، يمكن القول «إن الاختبار يمتاز

بخصائص سيكومترية جيدة، وإن جودته تزيد بعد مراجعة الفقرات التي تم تحديدها في العرض السابق، وبالتالي يمكن الاعتماد على هذا الاختبار في الحصول على بيانات دقيقة عن مستوى طالب الجامعة في المهارات الكمية.

كيفية إضافات فقرات جديدة للتطبيق المستقبلي:

أشرنا فيما سبق أن اختبار المهارات الكمية هذا سيستخدم لتقويم مهارات محددة لدى طلبة الجامعة وبشكل دوري، مما يؤدي إلى استهلاك فقراته، لذا فهناك حاجة إلى توفير آلية منتظمة يتم من خلالها تزويد هذا الاختبار بفقرات جديدة تماثل من حيث صعوبتها وتمييزها باقي الفقرات المستخدمة أصلاً في الاختبار. ولتحقيق ذلك، تم تبني طريقة مشهورة في بناء بنوك الأسئلة ومعادلة الاختبارات ضمن إطار نظرية الاستجابة على الفقرة، وهذه الطريقة هي تصميم الارتكاز (Anchor Design). ويقوم هذا التصميم حسبما أوضحه (Hambleton et al., 1991) على اختيار بعض الفقرات من الاختبار نفسه (تسمى فقرات الارتكاز) وضمها إلى الفقرات الجديدة المراد إضافتها لتكوين اختبار جديد، ثم يطبق هذا الاختبار على مجموعة مماثلة من الطلبة، وتستخدم خصائص فقرات الارتكاز لمعادلة خصائص الفقرات الجديدة ومن ثم إضافتها إلى الفقرات الموجودة أصلاً في الاختبار، وفيما يلي عرض للخطوات التي يتم بها ذلك:

١ - اختيار فقرات الارتكاز من ضمن الفقرات الموجودة في الاختبار الأصلي (لنسمي هذا الاختبار Y)، ويراعى في اختيار هذه الفقرات أن تمثل محتوى الاختبار، وأن تماثل في درجة صعوبتها ودرجة تمييزها بقية فقراته.

٢ - إعداد أو بناء الفقرات الجديدة التي سيتم إضافتها، وبالطبع يُراعى أن ترتبط كل فقرة جديدة بواحد أو أكثر من مخرجات الاختبار، وأن تماثل الفقرة في شكلها وصياغتها بقية فقراته.

٣ - إضافة فقرات الارتكاز إلى الفقرات الجديدة لتكوّن معاً اختباراً جديداً (لنسمي هذا الاختبار X)، ويتم تطبيق هذا الاختبار على عينة ممثلة من الطلبة بحيث يكون العدد كافياً لاستخدام النموذج الثلاثي المعلم

المستخدم في تحليل نتائج الاختبار. وبعد أن يطبق الاختبار (X)، يتم تحليل النتائج باستخدام أحد برامج الكمبيوتر في إطار نظرية الاستجابة على الفقرة وذلك لاستخراج قيم الصعوبة والتمييز لفقراته.

٤ - حساب المتوسط الحسابي ($\bar{b}_{Yc}$) والانحراف المعياري (S_{Yc}) لقيم الصعوبة لفقرات الارتكاز بناء على قيمها في التطبيق الأول (Y).

٥ - حساب المتوسط الحسابي ($\bar{b}_{Xc}$) والانحراف المعياري (S_{Yc}) لقيم الصعوبة لفقرات الارتكاز بناء على قيمها في التطبيق الجديد (X).

٦ - حساب قيمة الثابتين α و β حسب المعادلتين التاليتين:

$$\alpha = \frac{S_{Yc}}{S_{Yc}}$$

$$\beta = \bar{b}_{Yc} - (\alpha)\, \bar{b}_{Xc}$$

٧ - معادلة قيم الصعوبة (X_c) للفقرات الجديدة باستخدام المعادلة التالية:

$$b_{Yc} = \alpha(b_{xc}) + \beta$$

٨ - معادلة قيم التمييز (a_{Xc}) للفقرات الجديدة باستخدام المعادلة التالية:

$$a_{Yc} = \frac{a_{Xc}}{\alpha}$$

وبهذا يتم إضافة الفقرات الجديدة إلى الاختبار الأصلي وتحديد قيم الصعوبة والتمييز للفقرات الجديدة لتناسب بقية الفقرات الموجودة أصلاً في الاختبار وبالتالي إضافتها إليه، أما فقرات الارتكاز ذاتها فقد أصبح لكل منها قيم صعوبة وتمييز تختلف عن قيمها في الاختبار الأصلي (Y)، ويتم إعادتها إليه بقيم صعوبة وتمييز جديدة هي المتوسطات الحسابية لقيم صعوبتها وتمييزها في الاختبارين.

خاتمة:

تعتبر المهارات الكمية مكوناً أساسياً من مكونات برنامج التعليم الجامعي العام، ويجب أن يتم تقويمها بشكل دوري من خلال تطبيق أدوات قياس مناسبة ودقيقة، كما أن نتائج كل تطبيق يجب أن تتم دراستها ومراجعتها ومقارنتها بمعايير مستوى الأداء المتوقع سواء كانت محلية أو عالمية - وهي معايير

تحددها كل جامعة حسب سياستها التربوية - أو أن تقارن بنتائج تطبيق الاختبار في المرات السابقة لتحديد مدى فعالية الإجراءات والقرارات التي تتخذها إدارة الجامعة في سبيل تحسين مستوى مهارات الطالب الكمية.

وهدفت هذه الدراسة إلى تفعيل الإطار النظري في بناء أدوات قياس مقننة على المستوى الجامعي تعتمد في إعدادها وبنائها على الأهداف التربوية والمخرجات التعليمية أكثر من اعتمادها على المحتوى. وتعود أهمية مثل أدوات القياس هذه إلى أهمية البرامج المقوّمة من خلالها، ومنها برامج التعليم الجامعي العام الذي يشكل الأساس في التعليم الجامعي المشترك. ونظراً لعدم وجود محتوى محدد لمكوّن المهارات الكمية في هذا البرنامج، فقد تم بناء الاختبار باستخدام أهدافه العامة والمخرجات التعليمية التي يقيسها. أما بالنسبة لاستخدام نظرية الاستجابة على الفقرة كإطار لتحليل فقرات الاختبار واستجابات الطلبة، فقد ضمن ذلك الحصول على فقرات مناسبة ذات خصائص سيكومترية غير مرتبطة بمجموعة الممتحنين، ولكنها قادرة على قياس مدى تحصيل الطلبة على كل مخرج من مخرجات الاختبار، وهو ما لا يتوافر في النظرية الكلاسيكية في القياس. إضافة إلى ذلك، فإنه يمكن في إطار هذه النظرية بناء بنك أسئلة لاختبار المهارات الكمية يتوافر فيه ما يكفي من الفقرات ذات خصائص سيكومترية محددة.

كذلك فقد تم عرض طريقة لإضافة فقرات جديدة تعتبر فعالة للتزويد المستمر للاختبار بفقرات ذات خصائص سيكومترية مماثلة لما يتوافر في الاختبار الأصلي من فقرات، وهذا يعني ثباتاً في خصائص الاختبار ومستواه، كما أن ذلك يحول دون استهلاك فقراته بالاستعمال المتكرر. إن توافر مثل هذه الآلية يشكل إحدى مزايا استخدام نظرية الاستجابة على مستوى الفقرة في بناء وتحليل الاختبارات.

لقد هدفت هذه الدراسة إلى عرض المنهجية العلمية التي تم من خلالها بناء وتطوير اختبار لقياس المهارات الكمية في إطار نظرية الاستجابة على مستوى الفقرة، ويلاحظ من العرض أن هذه المنهجية قابلة للتطبيق في مجالات أخرى كثيرة كالقدرة اللغوية ومهارات التفكير وغيرها، بالإضافة إلى سهولة تطبيقها في الاختبارات التحصيلية.

Using Item Response Theory to Analyze Responses to a Quantitative Test for University Students

Dr. Hamzeh M. Dodeen
College of Humanities and Social Sciences
UAE University

Abstract

This study describes the methodology used by the United Arab Emirates University (UAEU) to develop a Quantitative Test at the university level. The expected educational outcomes were used to guide the process of developing the test items. The test was analyzed under the framework of the Item Response Theory (IRT) to periodically evaluate students' quantitative skills as one component of the University General Education Program (UGEP). The test consists of 40 multiple-choice items that cover five quantitative areas: algebra, statistics and probabilities, quantitative tables and graphs, geometry, and arithmetic. The methodology for developing the test included the following: defining the educational outcomes assessed by the test, piloting and reviewing the test intems, conducting the test on appropriate random sample (1005 students), determining the psychometric characteristics of the test, analyzing and evaluating the test and its items using IRT software, and using the Anchor Design mechanism for adding new items for future use of the test.

المراجع

1 - Bauman, S., & Martin, B. (1995). Assessing the quantitative skills of college juniors. **College Mathematics Journal, 26** (3), 214-220.

2 - Choi-Inn-Chul, & Bachman, F. (1992). An investigation into the adequacy of three IRT models for data from two EFL reading tests. **Language Testing, 9** (1), 51-78.

3 - Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. **Educational and Psychological Measurement, 58** 357-381.

4 - Hambleton, K., Swaminalton, H., & Rogers, H. (1999). **Fundamentals of item response theory.** Newbury Park, CA: Sage Publications.

5 - Hambur, S., & Jackson, D. (2000). Developing an instrument to assess the generic skills of students in Australian universities. Paper presented at the British Educational Research Association Conference, Cardiff University.

6 - Hendeson, M.; Morris, L.; & Fitz-Gibbon, C. (1987). **How to measure attitudes.** Newbury Park, CA: Sage Publications.

7 - Hockings, C. (2001). Evaluating alternative strategies for the development of mathematical thinking skills amongst undergraduate business studies students within the context of operations management. Paper presented at the British Educational Research Association Conference, Leeds.

8 - Linn, R., & Gronlund, N. (2000). **Measurement and assessment in teaching.** Columbus, OH: Merrill.

9 - Mislevy, R. J., & Bock, R. D. (1990). **BILOG maximum likelihood item analysis and test scoring with binary logistic models.** Mooresville, IN: Scientific Software.

10 - Oosterhof, A. (1994). **Classroom application of educational measurement.** NY: Macmillan College Publishing Company.

11 - Reckase, M.D. (1979). Unifactor latent trait models applied to multi-

factor tests: Results and implications. **Journal of Educational Statistics, 4**, 207-30.

12 - Richichi, R. (1996). **An analysis of test bank multiple-choice items using item response theory.** (ERIC Document Reproduction Service No. ED 405-367).

13 - Smith, R. (1996). A comparison of methods for detetrmining dimensionality in Rasch measurement. **Structural Education Modeling, 3** (1), 25-40.

14 - Stocking, M.(1990). Item response throry. In: G. Masters, & J. Keevs (Eds), **Advances in measurement in educational research and assessment.** The Boulevard, UK: Pengamon.