الفصل الثاني

بماذا يتميز اختبار اللغة المحوسب عن سواه؟
What is the CALT Difference?

انتهينا في الفصل السابق إلى أن المدرسين ومعدي الاختبارات والباحثين المعنيين باختبارات اللغة، بحاجة إلى إثراء معارفهم —كل في مجال تخصصه— من خلال اكتساب الجديد فيما يتصل بمدى التغيير الذي تحدثه التقنية في إمكانات التقييم اللغوي وواقعه. فما هي يا ترى تلك التغييرات المصاحبة الناتجة عن استخدام التقنية في التقييم اللغوي؟ ناقشنا التغييرات التي أحدثتها التقنية واستخداماتها في الجوانب ذات الصلة بمدرسي اللغة، ومعدي اختباراتها والباحثين المعنيين بهذه الاختبارات، غير أننا في هذا الفصل سنركز —تحديدًا— على كيفية اختلاف طرائق التقييم بمساعدة التقنية عن تلك الطرائق التقليدية القديمة التي يستخدم فيها معدو الاختبارات الورقة والقلم، ووسائط الصوت والصورة، والمقابلات الشخصية. وسنلقي الضوء كذلك على الجوانب الإيجابية للاستخدامات الحاسوبية، وسنتناول في عجالة بعض المشكلات المرتبطة بالتقنية المستخدمة في أعمال التقييم، وهو موضوع سنتناوله بشيء من التفصيل في الفصل الثالث، ونود التنبيه في هذا المقام على أن تحليلنا سيتناول المنظور الذي ثبتت جدواه لدى الباحثين في اختبارات اللغة فيما يتعلق بدراسة طرائق اختبارات اللغة. ومن هذا المنظور، سيكون من المفيد تناول طرق الاختبار من حيث الخصائص المميزة التي تؤلف

الحدث الاختباري بدلاً من تناولها بصفة شاملة في صورة مصطلحات، مثل: "الاختيار من متعدد" أو"المقابلة الشفهية" (Bachman, 1990). وسنستهل الفصل بموجز لخصائص الطريقة الاختبارية، ثم النظر بعد ذلك في الكيفيات التي تؤثر بها التقنية في كل واحدة من تلك الخصائص.

خصائص الطريقة الاختبارية

يتأثر استخدام اللغة بالسياق الذي يرد فيه هذا الاستخدام، ومن البديهي أننا نستخدم اللغة على صور شتّى. ويعتمد ذلك على: المكان، والشخص المُخاطَب، وسبب التواصل، وموضوعه، وشعورنا تجاه ذلك الموضوع، وما إذا كانت وسيلة التواصل هي الكلام أم الكتابة. غير أنه لا بد من سبر هذه الاختلافات بقدر أوفى من العناية والتدقيق عندما يكون "اختبار اللغة" هو السياق الذي تستخدم فيه اللغة. ومن المعلوم أن استخدام اللغة الذي يظهر على أداء الطالب، يتأثر بالخصائص السياقية المرتبطة بالمهمة قيد الأداء (الاختبار) مصحوبة بشرط غاية في الأهمية ألا هو العناية المبذولة في إدارة الاختبار في ظل موقف يحتم الاستخدام المصطنع للغة إن شئنا القول، إذ يخضع هذا الموقف لإدارة دقيقة في محاولة للتعاطي بإنصاف واتساق مع كل الطلاب في أثناء إجابتهم عن الاختبار. أما كون الاستخدام مصطنعًا، فراجع إلى حقيقة مفادها أن الاختبار يوضع للكشف بأقصى قدر ممكن من الصدق والأمانة عن جوانب محتملة من الاستخدام الفعلي للغة، بما يجعل ترجمة الأداء في الاختبار دليلاً على قدرة الطالب على الأداء اللغوي في واقع الحياة. ويعرف كارول، منذ نحو ٤٠ عامًا، الاختبار بأنه "إجراء مصمم لاستخلاص سلوك معين يمكن من خلاله استنباط أمور بشأن خصائص بعينها [تتعلق بالطالب] (Carroll, 1968، ص ٤٦). أما اليوم، فبإمكاننا

تعديل العبارة الأخيرة لتصبح كالتالي: "بشأن خصائص بعينها تتعلق بالطالب في سياقات غير اختبارية خاصة".

يصف الباحثون في اختبارات اللغة طرائق الاختبارات اليوم على نحو قريب مما سبق: فهي عندهم تتمثل في الطريقة التي نقدم بها الاختبار، والمهام المتنوعة التي نطلب إلى الطلاب تنفيذها، والوسائل التي نطلب إليهم الاستجابة من خلالها، إضافة إلى الإجراءات التي نستخدمها للحكم على مستويات أدائهم. وتؤثر هذه الخصائص على نحو يتعذر الفكاك منه عندما يستخدم الطلاب اللغة، وبذلك فهي تؤثر عند قياس قدراتهم اللغوية. ولما كانت التقنية تقدم صورًا جديدة لطرائق الاختبار، فمن المرجح أن يكون لذلك تأثير مهم في الأداء اللغوي للطلاب. ومن ثمَّ وجب تفهم مدى الاختلاف المحتمل للحواسيب عن الوسائل الأخرى في تقديم اختبارات اللغة ومهامها، وما يستتبع ذلك من مدى تأثير التقنية في تفسير الأداء في اختبارات اللغة. ويمكن أن يعكس ما يعرف باسم "أثر الطريقة" المرتبط بإجراءات تقديم الاختبار، ومعالجة الاستجابات ووضع الدرجات، الآثار الإيجابية والسلبية على الأداء في الاختبار. ومن المعلوم أن الآثار الإيجابية تتحقق عندما تعكس طرائقُ الاختبارِ خصائصَ اللغة المتعلمة ومواقف استخداماتها وجوانب استخدام اللغة التي وُضِع الاختبار لقياسها، كما أن الآثار الإيجابية تتحقق عندما تتوفر الإمكانات الحاسوبية المناسبة لتسجيل أو تحليل استجابات الطلاب على نحو يفوق إمكانات المُصححين من البشر. وعندما تبعد طرائق الاختبار كثيرًا عن استخدامات اللغة الواقعية، أو تتداخل مع قياس القدرات اللغوية المطلوبة، يتحول أثر الطريقة عندئذٍ إلى عامل سلبي. وفي كلتا الحالتين، يتعين على معدي الاختبارات والباحثين فيها، الإحاطة علمًا بالمدى المحتمل للآثار المترتبة على استخدام تقنية الحاسوب للإسهام في قياس القدرات اللغوية، كما يتعين عليهم الأخذ بما يمكّنهم من السيطرة على هذا المدى. ولمناقشة

اختلافات طرائق الاختبار، سنحذو حذو مؤلفين آخرين مشتركين في هذه السلسلة (مثلاً Alderson, 2000 وBuck, 2001 وDouglas, 2000) مستعينين بإطار عمل لهذه الطرائق يشمل الظروف المكانية والزمانية المحيطة بالاختبار، إلى جانب أمور أخرى مثل معايير أداء الاختبار ومدخلاته والاستجابات المتوقعة، والتفاعل بين المدخلات والاستجابات، وخصائص التقييم. وتجدر الإشارة إلى أن هذا الإطار مستنبط من الأطر التي وضعها باخمان (Bachman, 1990) وباخمان وبالمر (Bachman & Palmer, 1996) ودوقلاس (Douglas, 2000)، وكلها موضحة في الجدول رقم (٢,١). كما يتعرض الجدول ذاته لبحث المزايا والقيود الملموسة في اختبار اللغة المحوسب فيما يتعلق بكل خاصية من خصائص طريقة الاختبار.

وتشتمل الظروف المكانية والزمانية للاختبار على مكان إجراء الاختبار، والمسؤول عن تقديمه، وزمنه، علمًا بأن كلاً من باخمان وبالمر يستخدمان مصطلح "الخلفية" للتعبير عن هذه الخاصية، لكننا آثرنا مصطلح "الظروف المكانية والزمانية" لتمييزها عن المصطلح السابق عند نقاشنا لخصائص المدخلات والاستجابات فيما يلي. إن اختبارات اللغة المحوسبة —لا سيما إن كانت متاحة عبر شبكة الإنترنت— تريح الطلاب، حيث يكون بمقدورهم الجلوس لهذا الاختبار حيثما ووقتما أتيح لهم الوصول إلى الشبكة. غير أن هذه "الراحة" مقيّدة بقيود تفرضها مخاوف تتعلق بالحماية والسرية، لا سيما ما يرتبط منها بمسألة التأكد من هوية الطالب إذا كان الإنترنت هو وسيط الاختبار؛ فكلما كانت الاختبارات مصيرية للطلاب ومستخدمي نتائج الاختبار، زادت المخاوف ومباعث القلق. ومن القيود أيضًا (في المستقبل القريب على الأقل) مدى توافق الأجهزة والبرامج والخبرات المطلوبة للتثبيت والتركيب والصيانة، فكلها أمور تقيد الظروف المكانية الزمانية المثالية على أرض الواقع. أما مصطلح معايير الأداء، فيشير إلى

المعلومات المقدمة إلى الطالب بشأن كيفية الأداء في الاختبار. ومن ذلك: التعليمات، والزمن المخصص للإجابة، وتنظيم الاختبار، وكيفية الاستجابة للمهام المطلوبة فيه. ومن المسلم به أن الحواسيب توفر خاصية التوافق في تقديم التعليمات، وتخصيص زمن الإجابة، وتوحيد المطلوب في الاختبار، أضف إلى ذلك أنه يفسح المجال أمام عدد من الابتكارات كتوفير شاشات التعليمات (المساعدة) التي يمكن للطالب الاستفادة منها إن شاء، مما يزيد من حرية التصرف المتاحة له. وإذا كان النوع الأول قادرًا على تعزيز عنصر الاتساق والتناغم، فإن النوع الثاني قادر على زيادة عنصر التنوع مما يعني التأثير في قياس القدرات المطلوبة.

أما مصطلح "مدخلات الاختبار"، فيقصد به المادة المقدمة إلى الطالب ويُتوقع منه فهمها والاستجابة لها على نحو محدد عند أداء المهام المطلوبة في الاختبار. علمًا بأن المدخلات قد تتضمن طلبًا في حال إعداد مقال أو عند وجود مهمة لأداء دور بعينه على سبيل المثال، كما أن بعض المعلومات المرئية أو المسموعة كقراءة نص أو التعامل مع مخطط أو صورة أو مقطع فيديو تحتم على الطلاب التعامل معها عند استجاباتهم للأسئلة أو المهام المطلوبة بناءً على فهمهم لها.

وتؤثر تقنية الحاسوب كذلك على مدخلات الاختبار، وهو ما يؤثر بدوره في أسلوب الطالب واستجاباته على الاختبار؛ فالاستجابات المكتوبة عادةً ما تقدم بالكتابة أو بالنقر على الأزرار، بالإضافة إلى الكتابة باليد أو التحدث. ومن هذا المنطلق، نرى أن استخدام التقنية في إدخال استجابات الطلاب له ثماره البيّنة في الكشف عن مدى اعتياد الطلاب على الانخراط في استخدام اللغة بطريقة واقعية كما هو الحال في اللغة المتعلمة، أو مدى حدوث المشكلات في الحالات التي لا تتيح فيها التقنية استخدام أنماط استجابة تبدو للطلاب أنماطًا طبيعية معتادة. وما من شك في أن اعتياد الطلاب يتوقف على الخلفية المعرفية والتعليمية لكل فرد على حدة فيما يتصل باستخدام اللغة.

الجدول رقم (٢,١). خصائص الطرائق الاختبارية والمزايا والقيود في اختبارات اللغة المحوسبة.

خصائص طريقة الاختبار	مزايا اختبار اللغة المحوسب	قيود اختبار اللغة المحوسب
الظروف المكانية والزمانية: المكان والزمن والفرد	إمكانية أداء اختبار اللغة المحوسب في الكثير من الأماكن والأوقات المريحة، ودون تدخل بشري في الغالب.	مخاوف الحماية والأمان في الاختبارات المصيرية، ومن ذلك عدم توافق الأجهزة وعدم توافرها في كل مكان أصلاً، وكذا مقدار الخبرة التقنية المطلوبة للتثبيت والتركيب والصيانة.
معايير الأداء/التعليمات إجراءات الاستجابة	طلب مهام الاختبار على نحو يلائم جميع الطلاب، وتقديم التعليمات والمدخلات تلقائيًا وبصورة موحدة، مما يعطي فرصا متكافئة لجميع الطلاب.	اختلاف مستويات التعليمات ووجود شاشات اختيارية للمساعدة، واختلاف لغات التعليمات من شأنه الابتعاد عن عنصر التوحيد المطلوب.
المدخلات والاستجابات المتوقعة خواص السياق: الخلفية والمشاركون والأسلوب التنسيق: صورة/صوت/فيديو	قدرة الوسائط المتعددة على تنويع المدخلات وأنماط الاستجابة، مما يعزز من عنصر مراعاة السياق والاستيثاق من مختلف الجوانب.	تقييد عنصري المدخلات وأنماط الإجابة؛ نظرًا للتقنية المتاحة حاليًا.
التفاعل بين المدخلات والاستجابة التفاعل: تبادلي	قدرة الحواسيب على تكييف المدخلات وفقًا لاستجابات الطلاب وأدائهم، مما يتيح الاستفادة من الاختبارات الحاسوبية المتكيفة وعنصر التعقيب السريع.	القدرة على تحكم أفضل في التفاعل مقارنة بأنماط الاختبارات الأخرى، مما يعني تقييد قدرة الحاسوب على انتقاء العينات بحيادية وإنصاف، علاوة على التكلفة الباهظة لنشر الاختبارات المهيأة بالحاسوب.
خصائص التقييم تحديد المكونات معايير الصواب إجراءات التصحيح	قدرة معالجة اللغات الطبيعية على تقويم الاستجابات المعقدة تلقائيًا، مما يؤثر في تحديد المكونات ومعايير التصحيح وإجراءاته.	حداثة تقنية معالجة اللغات الطبيعية، وتكلفتها العالية وأوجه التقييد فيها، مما يؤدي إلى مشكلات محتملة حيال تحديد المكونات والصدق.

ولعل الجانب الأهم لعنصر "التفاعل بين المدخلات والاستجابة" – الذي يعنينا فيما يتعلق باختبارات اللغة المحوسبة – هو قدرة الحاسوب على تغيير المدخلات اللاحقة في ضوء استجابات كل مختبر على حدة، ويكثر هذا الأمر في الاختبارات الحاسوبية المتكيفة. (انظر: مناقشة الاختبارات الحاسوبية المتكيفة في الفصل الأول، ص ١١ – ١٣). يختلف التفاعل التبادلي القائم في إطار الاختبارات الحاسوبية المتكيفة اختلافًا كيفيًا عن التفاعل القائم في المحادثات التي تدور بين بني البشر. ولذلك فإن خاصية تكييف الاختبار التي يكفلها الحاسوب، تتيح خيارات جديدة للتفاعل من شأنها تفصيل الاختبارات لتناسب مستوى كل فرد من الأفراد.

وأخيرًا، يشير مصطلح "خصائص التقييم" إلى تحديد ما ينبغي قياسه بالاختبار، والمحكات التي يجري على أساسها قياس الأداء وفقًا لما سبق تحديده، إلى جانب العملية التي يقيم من خلالها الأداء أو وضع الدرجات. من جانب آخر، تتزايد الأساليب المعقدة التي يسهم بها الحاسوب الآلي في التصحيح من خلال تقنية معالجة اللغات الطبيعية (NLP) التي تحاكي بعض جوانب الحكم البشري على الأمور. بيد أن هذه التقنية ما تزال في مهدها، ومكلفة نسبيًا ومحدودة بقيود تتعلق بجوانب اللغة الشفهية والمكتوبة التي يمكن قياسها، الأمر الذي يتطلب البحث والتنقيب وصولاً إلى أفضل الاستخدامات للإمكانات الحالية لتحليل الاستجابة.

نخلص مما سبق إلى أن خصائص الطريقة الاختبارية، توفر إطار عمل ينظم تحليل أوجه الاختلاف والتميز التي تتسم بها اختبارات اللغة المحوسبة مقارنة بما سواها؛ نظرًا لأنها تعكس نقاط اتخاذ القرار التي يظهر فيها تأثير معدي الاختبارات على ما وُضع الاختبار لقياسه من خلال تحديد كيفية القياس الذي يضطلع به الاختبار، علمًا بأنه لا تتوافر سوى بحوث قليلة تكشف عن الآثار المترتبة على خيارات

التقنية في هذا الصدد. ومن ثمّ يهدف هذا الفصل إلى استجلاء النقاط المهمة التي تكتنف التداخل بين التقنية وأوجه الطرائق الاختبارية من خلال الوصف والتفسير، فيما يتناول الفصل التالي بعض النقاط الخلافية التي وجه إليها الكثير من النقد بشأن استخدام التقنية في اختبارات اللغة.

الظروف المكانية والزمانية للاختبار

لطالما صحبت اختبارات اللغة مشكلاتٌ، مثل: إمكانية وصول الطلاب إلى مقر الاختبار ويسر التطبيق والإجراء، والنفقات المرتبطة بالاختبار؛ فشروط إجراء الاختبار في معظم الاختبارات التقليدية، تنص على أن يتقدم الطلاب لأداء الاختبار في مكان معين ووقت محدد. أما برامج الاختبار التي تجرى على نطاق واسع(مثل: "النظام العالمي لاختبار اللغة الإنجليزية" (IELTS) وبطارية ميتشجان لتقييم اللغة الإنجليزية (MELAB) أو الامتحانات القومية مثل اختبار اللغة الإنجليزية بالكلية الوطنية في الصين) فتستوجب كلها إرسال مواد الاختبار إلى مقار الاختبار وسفر الطلاب مسافات بعيدة في بعض الأحيان لأداء الاختبار، ويلي ذلك إعادة استجاباتهم إلى مراكز التصحيح، ثم إرسال النتائج إلى الطلاب ومستخدمي نتائج الاختبار. وحتى الاختبارات التي تجرى على نطاق ضيق نسبيًا(مثل:اختبارات تحديد المستوى أو اختبارات التحصيل التي تجرى في الصفوف الدراسية) يجب على الطلاب الوصول في زمان ومكان محددين، حيث يوزع مسؤول الاختبار أو المعلم أوراق الاختبار، ويقرأ التعليمات بصوت عال ويتعامل مع الأجهزة المتاحة –كمشغلات الأشرطة الصوتية– ويراقب إجراءات الاختبار، ويراعي الزمن المتاح له، ثم يجمع أوراق الاختبار ومواده إيذانًا بانصراف الطلاب؛ أي أن تطبيق أي اختبارات لغوية وإجرائها يتطلب في الغالب التزامًا حقيقيًا بعنصري الزمن والمجهود.

أما الاختبارات المحوسبة، فتقلل من أعباء التطبيق والإجراءات المتمثلة في المراقبة أو الإشراف من خلال إتاحة الاختبار حيثما ووقتما تسنى للطالب الوصول إلى الإنترنت أو إدخال أسطوانة في محرك الأقراص. وبالمثل، يمكن للاختبارات المحوسبة تقليل الأعباء الإدارية من خلال نقل مواد الاختبار إلكترونيًا، مما يتطلب عددًا أقل من المسؤولين عن توزيع أوراق الاختبار وجمع أوراق الإجابة، وتشغيل المعدات المتاحة وما إلى ذلك. كما نجد أن الاستعانة بشبكة الإنترنت في تقديم الاختبارات يشكل أهمية خاصة لمعدي الاختبارات من حيث إمكانية استغلال التقنية في تحسين إمكانية الوصول إلى الشبكة، ولعل تزايد معدلات الوصول هو الفائدة الأعظم التي يؤكد عليها أنصار الاختبارات عبر الويب، ومن بينهم روفر (Roever, 2001) إذ يقول:

> لعل أعظم ميزة لوجيستية للاختبارات عبر الويب (WBET) هي مرونة الاختبار زمانًا ومكانًا، فكل المطلوب لأدائها هو توفر حاسب آلي به برنامج لتصفح الإنترنت والاتصال بها (أو أداء الاختبار من خلال اسطوانة). ويمكن للطلاب أداء هذا النوع من الاختبارات وقتما وحيثما تيسر لهم ذلك، فيما يمكن لمصممي الاختبار تبادل الاختبار مع زملائهم في كل أنحاء العالم وتلقي التغذية الراجعة عليه (ص ٨٨).

من الأمثلة على الاختبارات التي تجرى عبر الإنترنت، والمصممة لخدمة قطاع عريض من المستخدمين وتسهيل وصولهم إليها، ذلك الاختبار الذي أعد ليكون بمثابة وسيلة للمواطنين الأوروبيين لتقييم مهاراتهم ومعارفهم، والاعتراف بها بعيدًا عن أنظمة التأهيل الرسمية المعمول بها في بلادهم (DIALANG, 2001). ويتيح هذا الاختبار للمستخدمين اختيار اللغة التي يفضلونها لتلقي تعليمات الاختبار بها وكذا التغذية الراجعة، وكذلك اختيار المجالات اللغوية والمهارية التي يريدون الاختبار فيها (كالقراءة والكتابة والاستماع والتراكيب النحوية والمفردات). وقد أدرك مشروع

دايلانق (DIALANG) الفوائد المتحققة من اختبارات اللغة المحوسبة للطلاب والمسؤولين عن الاختبارات على حد سواء، والتي تتمثل في سهولة الوصول إلى الاختبار وتقديمه عبر الإنترنت. ويمكن للمستخدمين الوصول إلى هذا الاختبار غير المصيري في ظروف مناسبة لهم تمامًا وقتما وحيثما تسنى لهم الوصول إلى شبكة الويب، بل إنه يوفر للطلاب فرصة تقييم كتابتهم ذاتياً بأنفسهم مستعينين في ذلك بمجموعة من المعايير (Luoma and Tarnanen 2003). ومن وجهة نظر المسؤولين عن الاختبار، يتميز الاختبار بأنه قابل للإدارة الذاتية بصورة كاملة، حيث يقتصر دور المسؤولين عنه على جمع البيانات واستكشاف الأعطال وإصلاحها. وعندما يطبق الاختبار بشكل كامل، يمكن للمسؤولين عنه جمع البيانات المطلوبة عن الأداء في كل بند من بنوده، والوقوف على الزمن الذي يستغرقه الطلاب للإجابة عن كل من هذه العناصر، وكذلك جمع بيانات عن الاختبار عموماً وعن الاختبارات الفرعية التي تندرج تحته،بما في ذلك التقييم الذاتي وإمكانية إعادة الاختبار، ومعرفة خصائص المستخدمين الديموغرافية. ومن شأن جمع المعلومات بهذه الطريقة الروتينية، تمكين القائمين على الاختبارات من إجراء مراجعات مستمرة، علمًا بأنه يمكن الاستفادة من مزايا الويب على نحو أكبر عندما تكون الاختبارات غير مصيرية. أما في حالة الاختبارات المصيرية، فإن خيارات كالوصول الذاتي للاختبار، وتقييم الطلاب لأنفسهم ذاتياً، تقتضي أن نتعامل معها بصورة مختلفة عن تلك التي نتعامل بها مع الاختبارات غير المصيرية.

معايير الأداء والتعليمات

هناك قلق متزايد باطراد فيما يتعلق بتباين أسلوب إجراء اختبارات اللغة التقليدية من اختبار إلى آخر من حيث تقديم التعليمات، وكيفية التحكم في وقت كل جزء من أجزاء الاختبار ومتابعته ؛ إذ يمكن أن تقدم التعليمات مقروءة أو أن يقدمها

المسؤولون عن الاختبار داخل قاعات الاختبار الكبيرة ارتجالاً في بعض الأحيان حيث تتباين الخصائص الصوتية للإلقاء تباينًا شديدًا للطلاب الجالسين في أماكن قد تكون بعيدة أو قريبة من مصدر الصوت. كما أن المحتوى الفعلي للتعليمات، يختلف باختلاف خبرة المسؤول وطريقة تفكيره، أضف إلى ذلك أن تحديد زمن معين لكل جزء أو مهمة من مهام الاختبار يمثل بدوره مصدرًا محتملاً للتباين بين الإداريين المشرفين على الاختبار. وعلى النقيض من ذلك، نجد أن توجيهات الاختبار المحوسب تُقدم بأسلوب واحد لكل الطلاب، حيث إن الزمن المخصص لكل جزء أو مهمة في الاختبار محدد تلقائيًا وعلى نحو يمكن الوثوق به. كما أن الحواسيب توفر في الوقت ذاته مرونة كافية إزاء كيفية إصدار التعليمات ووقتها وباللغة المناسبة. فهناك مثلاً مركز "خدمة الاختبار اللغوي لرجال الأعمال" (المعروفة اختصارًا بالإنجليزية BULATS) التي تتيح للطالب تلقي التعليمات بإحدى ست لغات أوروبية.

ثمة جانب آخر لمعايير الأداء والتعليمات المقدمة في الاختبارات التدريبية والتحضيرية التي تسبق الاختبار الفعلي، ففي معظم الاختبارات التدريبية غير المحوسبة عادةً ما تكون المواد المستخدمة منفصلة تمامًا عن تلك المقدمة في الاختبار الفعلي. وعلى الرغم من تشابهها مع المواد المستخدمة في الاختبار الفعلي؛ من حيث الشكل والمكان الذي يعقد فيه الاختبار، فإنها تتباين في جوانب أخرى. ولنا أن نشير إلى اختلاف رئيس هو سهولة اختبارات اللغة المحوسبة، وإتاحة المواد التحضيرية والتدريبية للاختبار على نطاق واسع بما يتيح التدرب عليها في ظروف مشابهه للاختبار الفعلي، ذلك أن المواد المنشورة على شبكة الإنترنت، أو على أقراص مدمجة، تتيح للطلاب فرصة التدرب على أنماط وطرق الإجابات المختلفة، وهو ما يماثل ما يجده الطلاب في الاختبار الفعلي. ويعدُّ موقع اختبار "التوفل" (http://www.ets.org/toefl) أفضل مثال على ذلك، حيث يتيح عددًا من الخيارات التدريبية، بما في ذلك إمكانية التدريب على الاختبار مجانًا عبر الإنترنت.

وخلاصة القول أن توفير معايير لأداء الاختبار وكذلك التعليمات الخاصة به؛ بالإضافة إلى سهولة الوصول إلى المواد التحضيرية، تمثل كلها اختلافًا مهمًا يصب في صالح اختبارات اللغة المحوسبة. ولما كان كل الطلاب يتلقون المدخلات نفسها بدقة، بغض النظر عن مكان الاختبار وزمانه، ولما كان الوصول إلى المواد التحضيرية والتدريبية متاحًا لكل الطلاب عبر الوسيلة ذاتها التي يُقدم الاختبار من خلالها، فإن هذه العوامل تتضافر لتولد شعورًا لدى الطلاب كافة بفرصهم المتكافئة للنجاح. وفي الوقت ذاته، فإن أنظمة تقديم اختبارات اللغة المحوسبة وإجراءاتها المعقدة، قد تخلق حاجة إلى توفير مواد تدريبية خاصة تناسب هذا النوع من الاختبارات أكثر من الأنواع الأخرى.

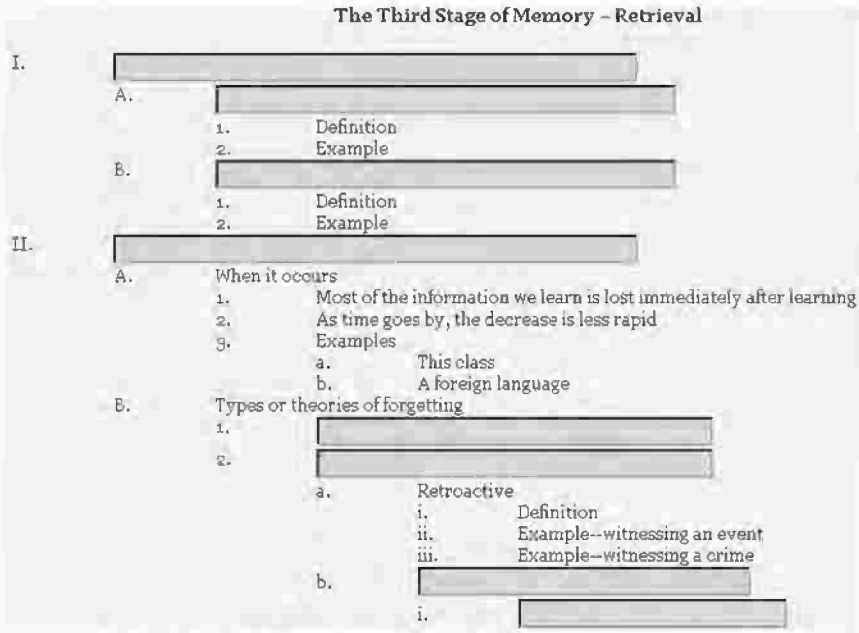
مدخلات الاختبار والاستجابات المتوقعة

يمكن استخدام الحواسيب لعرض كم ضخم من البيانات ومعالجتها على نحو سريع، مما يتيح تضمين معلومات سياقية ثرية في المدخلات التي تقدم لمختبري اللغة، ومن ذلك: الصور، والأصوات، ومقاطع الفيديو الحركية، الأمر الذي من شأنه تعزيز عنصر الأصالة في المدخلات والاستجابة على حد سواء. وهذه الإمكانية تفسح المجال أمام التعاطي مع مشكلة قائمة في اختبارات اللغة وهي أن الاختبارات عادة ما يُنظر إليها بأنها مكونة من مدخلات ومهام منزوعة تمامًا عن السياق اللغوي الذي تستخدم فيه اللغة. وهذه مشكلة تؤثر في عنصر الأصالة؛ نظرًا لأن استخدام اللغة في الواقع الذي نعيشه، يكون مضمنا" في السياقات الاجتماعية والنفسية، في حين تشتمل كثير من الاختبارات التقليدية على لغة مصطنعة تستلزم إعادة صياغتها بصورة أو بأخرى لتناسب السياق الذي تستخدم فيه. والسؤال الذي يطرح نفسه الآن هو: كيف يمكن أن نصل بالجانب السياقي إلى أقصى درجة من الفعالية؟ فعلى الرغم من أن اختبارات اللغة غير المحوسبة قد وظفت بعض الأساليب السياقية الفعالة، فإنها تظل

مقتصرة عموماً على ما يمكن تحقيقه من خلال الوسائل المطبوعة مصحوبة بوسائط الصوت والصورة، وعلى المقابلات المباشرة وأداء الأدوار التي تعتمد على الجانب الإبداعي والقدرة لدى الأشخاص الذين يقومون بتلك المقابلات. ومن ثمّ نجد أن تقديم المهام الاختبارية التي تعكس تنوع اللغة وتعقيداتها لاستخدامها في أغراض التواصل، يشكل تحديًا دائمًا أمام معدي اختبارات اللغة.

تقدم اختبارات اللغة المحوسبة لمعدي الاختبارات خيارات جديدة لتقديم المدخلات على وسائط متنوعة، بما في ذلك وسائط النصوص، والرسوم، والصوت، والفيديو إلى جانب تحكم المستخدم في المدخلات. وبالقدر الذي تعكس به المدخلات متعددة الوسائط خصائصَ اللغة الهدف ومواقف استخداماتها المختلفة، فإنها تساعد أيضا على تعزيز عنصر الأصالة، وزيادة الاهتمام بمهام الاختبار ذاتها؛ مما يعزز احتمال التفاعل على نحو أكبر بين قدرة الطالب اللغوية التواصلية من جانب، ومهام الاختبار من جانب آخر. ولقد تناولنا في الفصل الأول مثالاً من اختبار "نظام التقييم اللغوي عبر الانترنت" (WebLAS) بجامعة كاليفورنيا في لوس أنجلوس، ورأينا كيف اشتمل الجزء الخاص بالاستماع في هذا الاختبار على نص من نصوص الفيديو يستغرق ست دقائق ونصف تقريبًا استقي من محاضرة في علم النفس تتناول موضوع استعادة المعلومات من الذاكرة (UCLA, 2001). وفي أثناء هذا الاختبار يشاهد الطلاب مقطع فيديو على شاشة صغيرة يتحدث فيه أحد المحاضرين مستخدماً جهاز عرض لكتابة ملاحظات موجزة على الشاشة، ويقوم بحركات مصاحبة في أثناء المحاضرة: كالتحرك، واستخدام بعض الإيماءات، وتعبيرات الوجه وما إلى ذلك. ويرى معظم معدي الاختبارات في هذا النمط درجة عالية من الأصالة وتمثيل الواقع للمحاضرة الحقيقية، ويرجع هذا إلى قدرة الطالب على رؤية أشياء مرتبطة بقاعة المحاضرات مثل جهاز العرض والشاشة والسبورة، فضلاً عن سماعه لمحاضر يلقي محاضرة أكاديمية

رسمية واقعية تتناول موضوعًا أكاديميًا مستعينًا بمصطلحات فنية وشبه فنية ومفردات في صميم تخصص علم النفس. وبناء على ذلك نجد أن الاستجابة المطلوبة –كما صورت في الشكل رقم (٢.١)– عبارة عن توضيح لملاحظات عن تلك المحاضرة التي يُنظر إليها باعتبارها مهمة أكاديمية مستوفية الشروط. وما ذلك إلا مثال واحد على إمكانية استخدام تقنية الحاسوب لتوفير مدخلات صوتية ومرئية ثرية للطلاب، في محاولة لانخراطهم في إجراء تقييمي في موقف يتسم بالتفاعل والأصالة والقرب من الواقع.

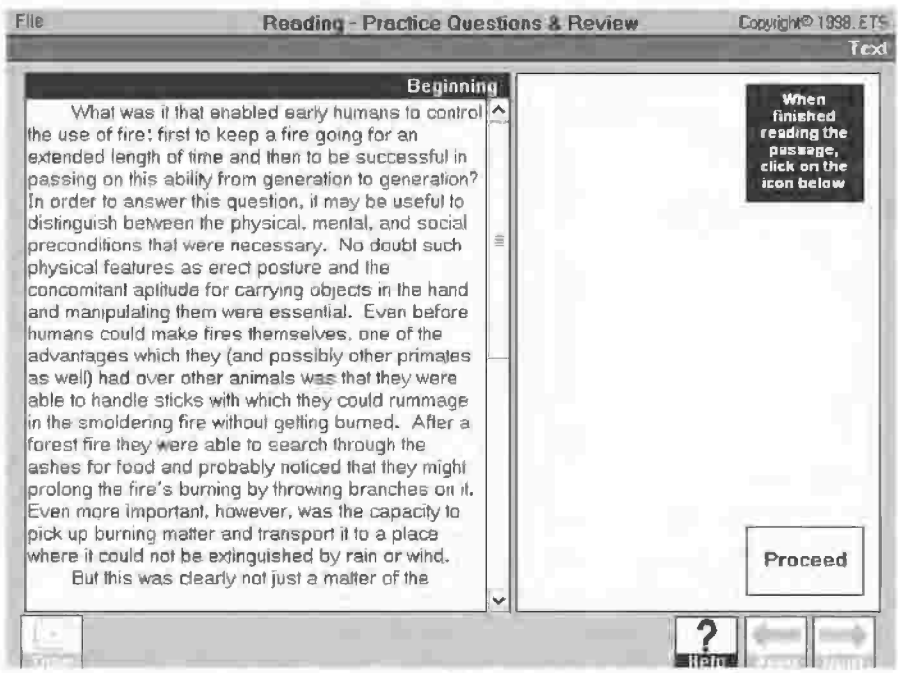


الشكل رقم (٢,١). لقطة من نظام التقييم اللغوي عبر الإنترنت لمهمة استماعية.

(http://www.humnet.ucla.edu/web/departments/alt/weblas_esl_demo/demo_listen_out_psych1.htm)

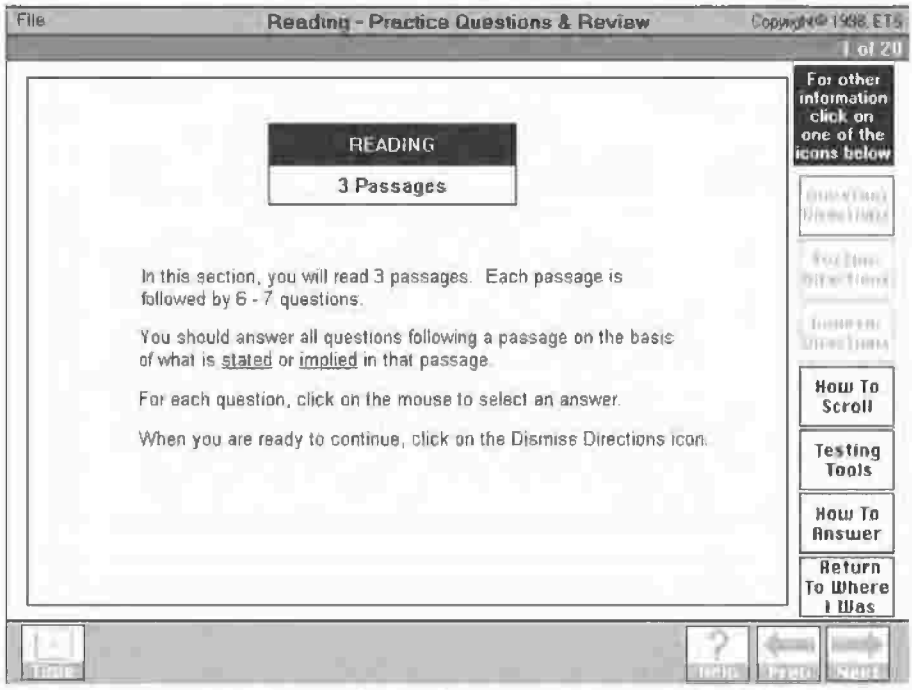
تفتح التقنية المجال أيضاً أمام خصائص جديدة تتعلق بالاختبارات، ومن ذلك إمكانية الوصول الفوري للمصادر المساعدة كمسارد المصطلحات، أو المعاجم، أو الطرق الإرشادية في أثناء الاستجابات التي يقوم بها الطالب. فعلى سبيل المثال، يحتوي

القسم الخاص بالقراءة في اختبار "التوفل" المحوسب (CBT) كما هو موضح في الشكل رقم (٢,٢) على زر للمساعدة (Help) ويوجد بالجزء السفلي الأيمن للشاشة، ويمكن للطالب الضغط عليه لمشاهدة شاشة تعرض توجيهات خاصة بالقسم على نحو متكرر، وكذلك توجيهات تتعلق بالطرائق الملائمة لأداء الاختبار، بما في ذلك كيفية التنقل بين أقسام الاختبار وعناصر التحكم فيه مثل خيارات "التالي" (Next) و"التأكيد" (Confirm) و"الزمن" (Time) والتحكم في "مستوى الصوت" (Volume)، بالإضافة إلى بيان كيفية الإجابة عن كل نوع من المهام المطلوبة كما هو موضح في الشكل رقم (٢,٣).



الشكل رقم (٢,٢). لقطة من اختبار التوفل المحوسب الخاص بقسم القراءة.

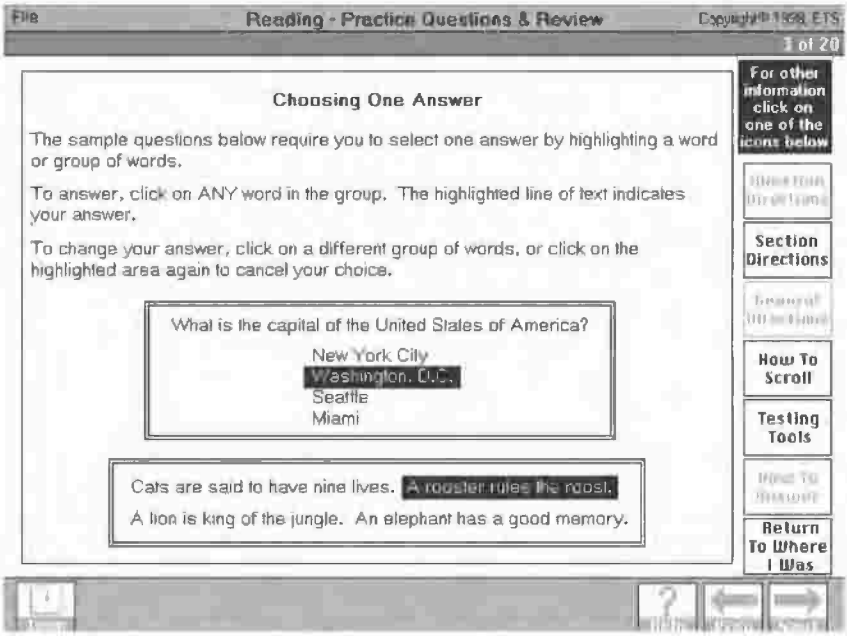
(Educational Testing Services, 1999)



الشكل رقم (٢,٣). لقطة من شاشة التعليمات في اختبار التوفل المحوسب.

(Educational Testing Services, 1999)

أما زر "كيف تجيب؟" (How to Answer) فيأخذ الطالب إلى شاشة معينة كما هو موضح في الشكل رقم (٢,٤)ويقدم معلومات حول كيفية مباشرة الطلاب للمهام التي تتطلب تظليل أجزاء من النص على الشاشة. ووفقا لما يراه معدو الاختبارات، فإن تمكن الطلاب من الوصول الفوري للتوجيهات الاختيارية والمعلومات التي تساعدهم على الاستجابة تعتبر من المزايا المفيدة لبعض الطلاب . ولعل هذا المثال أيضا يقدم لنا نوعًا جديدًا من المهام الاختبارية التي توفرها التقنية.



الشكل رقم (٢,٤). لقطة من اختبار التوفل المحوسب الذي تظلل فيه الاستجابات.
(Educational Testing Services, 1999)

إن هذه الإمكانات والقدرات المتمثلة في مراعاة السياق، وتوفير لخصائص المساعدة، وأنواع المهام الجديدة، لم يسبر أغوارها ولم يفهمها معدو الاختبارات ولم يستخدموها على النحو الأمثل بعد ؛ فهذا "فولتشر" (Fulcher, 2000) يوضح بأن معدي الاختبارات يساورهم بعض القلق الذي له ما يبرره حيال استخدام الجوانب التفاعلية للوسائط المتعددة والحاسوب، وما قد يتمخض عن ذلك من تأثير في الجوانب المراد قياسها في الاختبار بسبب "أثر الطريقة" التي ذكرناها سابقاً". ولذلك فإن معدي الاختبارات لا يودون أن يتسرعوا في توظيف خيارات اختبارات اللغة المحوسبة توظيفاً كاملاً إلا بعد المعرفة التامة لأثر ذلك في أداء الطلاب، وتفسير نتائج اختباراتهم.

التفاعل بين المدخلات والاستجابات

يمكن أن تؤثر طرائق اختبار اللغة المحوسبة في التفاعل الذي يحدث بين مدخلات الاختبار واستجابات الطلاب فيما يتعلق بالاستجابات، وأسلوب التغيير في المدخلات، أو التفاعل معها في ضوء استجابات الطالب. ففي أشكال الاختبارات المختلفة ومن بينها —على سبيل المثال— المقابلات المباشرة التي تهدف إلى الكشف عن بعض الجوانب لدى الممتحن في أثناء المحادثة وجهًا لوجه، يكون أمام من يضطلع بالمقابلة الكثير من الخيارات لتعديل المدخلات: كزيادة مستوى صعوبة المفردات أو التراكيب اللغوية أو تيسيرها؛ لاسيما إذا كان في أداء الطالب —عند بداية الاختبار— ما يوحي بارتفاع أو انخفاض في مستوى قدراته. كما يمكن لمن يضطلع بالمقابلة تغيير الموضوع ليصبح أكثر توافقًا مع اهتمامات الطالب وأغراضه، فضلاً عن قدرته على أن يستثير بعض الاستخدامات اللغوية التي يحاول الطالب تجنبها. كما يمكن لمن يضطلع بالمقابلة من ذوي الخبرة، تلمُّس مستوى الطالب الحقيقي من خلال النظر في القدرات اللغوية العليا والدنيا له والتدقيق فيها أكثر من مرة، ومن ثم الاستقرار على النقطة التي تمثل المستوى المناسب للطالب الذي يمكن أن يحافظ فيه على مستوى أدائه (انظر مثلاً Liskin-Gasparro, 1984 للوقوف على وصف "المجلس الأمريكي لتعليم اللغات الأجنبية" (American Council on the Teaching of Foreign Languages) بشأن "المقابلات الشفهية لتحديد مستوى الكفاية في اللغة" (Oral Proficiency Interview).

هذا النوع من المرونة —أو التكيف— لا يمكن الحصول عليه في الاختبارات الورقية؛ كاختبارات القراءة، أو اختبارات الاستماع المسجلة على شرائط كاسيت حيث تُقدَّم المدخلات على نحو متتابع يوجب على الطلاب محاولة الإجابة عن كل المهام المطلوبة سواء أكانت غاية في السهولة أم بالغة التعقيد. ويبدو أن تقنية الحاسوب ستستغرق وقتًا طويلاً قبل أن تتمكن من محاكاة الإنسان الذي يتميز بالمرونة عند إجراء المقابلات الخاصة بالاختبارات الشفهية. غير أن اختبارات اللغة المحوسبة —كما ذكرنا في الفصل الأول— تتيح إلى حد ما درجة من التفاعل بين المدخلات والاستجابات في اختبارات الاستماع أو القراءة أو الاختبارات الخاصة بمعرفة التراكيب النحوية. وبناء على ذلك، فإن دقة الحاسوب وموثوقيته

في تقييم الاستجابات والقدرة على التكيف معها، قد تبلغ مستوى يفوق المستويات التي يمكن للأشخاص القائمين على الاختبارات تحقيقها على أرض الواقع.

وقد أصبح الاختبار الحاسوبي المتكيف أمراً "ممكناً" بفضل تطور نظرية الاختبار التي تخطت حدود نظرية التقييم التقليدية التي ترتكز عليها المفاهيم التقليدية في تقييم مستوى الثبات وحساب درجات كل بند من بنود الاختبار (كالوقوف على مدى صعوبة كل بند مثلاً). أما نظرية الاستجابة للبند (IRT) فتستند إليها إجراءات التحليل الكمي للبنود التي تنتج عنها إحصاءات مستقلة لكل بند بغية الوقوف على صعوبة البند وتحديد المعايير الأخرى ذات الصلة بقدرات محددة يُراد قياسها لدى الطالب. وبعد التحليل الأولي لبيانات ما قبل الاختبار، يتم "ترميز" كل بند بمستوى معين من الصعوبة توازي فرصة الإجابة عنه إجابةً صحيحة ما نسبته ٥٠٪ من الطلاب من العينة نفسها. من جانب آخر، يوظف عدد من اختبارات اللغة التي تطبق على نطاق واسع، مثل: اختبار "التوفل" والاختبار الأوروبي (DIALANG) الأساليب الخاصة بنظرية الاستجابة للبند والاختبارات الحاسوبية المتكيفة. أما الاختبارات التكيفية القائمة على أساليب النظرية المذكورة فتتناسب بصفة خاصة مع الاختبارات المؤلفة من مجموعة من البنود المستقلة، غير أنه قد استخدمت أنواع أخرى مختلفة لتقييم المهارات التي تتطلب مجموعة كبيرة نسبيًا من النصوص في صورة مدخلات؛ كمهارتي الاستماع والقراءة. وعلى الرغم من وجود أمثلة على اختبارات لمهارة القراءة تتألف من بند واحد لكل قطعة (مثل Choi, Kim, & Boo, 2003)، فإن مستوى التكيف للبنود لا يعتبر فاعلاً أو عملياً؛ لأنه ينبغي على الطالب القراءة لعدة دقائق للبحث عن كل بند يحتاج إليه في الاختبار. ومن ثمّ أعدت ما تعرف باسم "مجموعات النصوص والأسئلة" أو ما يعرف اختصارًا بالإنجليزية باسم "Testlets" – بحيث تمثل كل مجموعة مستوى مختلفًا من القدرات (Bradlow, Wainer & Wang, 1999). وفي أثناء الاختبار، يختار الحاسوب قطع قرائية كاملة مع أسئلتها بدلاً من اختيار بنود مفردة استنادًا إلى الاستجابات الصحيحة أو الخاطئة المبنية على بنود سابقة. فعلى سبيل المثال، يتألف اختبار الكلية الأمريكية المعروف اختصاراً باسم

(ACT) وهو اختبار لتحديد المستوى (كان يعرف سابقًا باسم "اختبار الكلية الأمريكية المسجل" (American College Test, Incorporated) من ثلاث وحدات هي: القواعد/ الاستخدام، والقراءة، والاستماع التي يمكن إجراؤها كل على حدة أو مجتمعة، في حين يقدم كل اختبار درجة منفصلة لتحديد المستوى. ويوضح الشكل رقم (٢,٥) مجموعات النصوص والأسئلة الخاصة بالقواعد/الاستخدام.

Level 4

Item Types: (1) Passive voice verbs; (2) Punctuation conventions: parentheticals; (3) Conjunctions; (4) Noun clauses; (5) Noun clauses

The lion **(1)** _____ the "king of beasts," but in fact, this king doesn't do much to deserve the title. Lions live in groups, each group, or "pride," of lions consists of one dominant male and several females and their offspring. However, all the **(2)** _____ is done by the lionesses, the females. Males spend their time **(3)** _____ what they have killed. Occasionally, males may have to scare off possible threats to the group, but for no more effort than this, they get to eat first, and to eat until they are full. Male lions are also guilty of what **(4)** _____ not very kingly behavior. When a new male takes over the pride, he will usually kill all the cubs of the male that formerly led the pride. This ensures that only his own offspring will survive. This may seem brutal, but it is how **(5)** _____. Perhaps we might want to reconsider calling the lion the "king of beasts."

1.
- ☐ A. is often been calling
- ☐ B. has often called
- ☐ C. is often been called
- ☑ D. has often been called

2.
- ☑ A. work—the stalking, chasing, and killing of prey—
- ☐ B. work—the stalking, chasing, and killing of prey
- ☐ C. work,—the stalking, chasing, and killing of prey,
- ☐ D. work —the stalking, chasing, and killing of prey

3.
- ☑ A. either sleeping while the females hunt or eating
- ☐ B. sleeping while either the females hunting or eat
- ☐ C. either sleeping while the females hunt or eat
- ☐ D. sleeping while the females either hunt or eating

4.
- ☐ A. would we probably call
- ☑ B. we would probably call
- ☐ C. we would probably call it
- ☐ D. would we probably called

5.
- ☐ A. nature does working
- ☐ B. is nature working
- ☑ C. nature works
- ☐ D. does nature work

الشكل رقم (٢,٥). اختبار تحديد مستوى القواعد.

وفي هذه المجموعة المأخوذة كمثال، يقدَّم النص والأسئلة المرتبطة به دفعة واحدة، في حين يحدد أداء الطالب في هذه المجموعة مستوى القدرة المطلوب في المجموعة التالية المقدمة في الاختبار. وفي ضوء الحاجة إلى تضمين خطاب متصل كمُدخَل في الاختبارات اللغوية، تتيح المجموعات المذكورة بُعدًا إضافيًا لاستخدام الاختبارات الحاسوبية المتكيفة لأغراض التقييم اللغوي.

وعلى الرغم من أن هذه الاختبارات لا توفر النوع نفسه من التكيف، مثلما هو الحال مع الأساليب التفاعلية المباشرة، كالمقابلات الشخصية على سبيل المثال، فإنها تتفوق كثيراً على الطرق التقليدية في قياس بعض القدرات اللغوية. لكنها تنطوي في الوقت ذاته على تحديات تواجه معدي الاختبارات، لاسيما في ضمان التغطية المناسبة لنطاق من المحتويات، وكذا تقديم مجموعة من المهام اللغوية على نحو يكفي لتمثيل القدرات المطلوب قياسها. ومن أجل التأكد من تغطية نطاق عريض من المحتوى ومستويات القدرات، تتطلب الاختبارات الحاسوبية المتكيفة إعداد كم هائل من البنود واختبارها مسبقًا، الأمر الذي يجعل من بناء الاختبار أمرًا مكلفًا للغاية يهدر الكثير من الوقت. كما يجب بناء اللوغاريثم التكيفي – أو القواعد التي ستحكم انتقاء الحاسوب لكل بند لتقديمه – بعناية فائقة لضمان حساسيته للعوامل اللغوية والموضوعية في المحتوى وكذا نوع المهمة، إلى جانب ما يتضمنه أصلاً من درجة الصعوبة في كل بند يستجيب له الطالب عند دخوله مستوى محدد من القدرات. من هنا نتوقع أن يشغل التحدي الماثل في إعداد اختبارات اللغة الحاسوبية المتكيفة (ذات القدرة على مراعاة تقييم القدرة اللغوية التواصلية) بال معدي اختبارات اللغة والباحثين في هذا المجال على مدار العقود المقبلة. فعلى سبيل المثال، قد يكون بمقدور الحواسيب في المستقبل استغلال المدخلات الكلامية على نحو تكيفي، وبذلك تتيح محاكاة خصائص التواصل بالكلام البشري على نحو لا يخطر على بال أحد في الوقت الحاضر.

خصائص التقييم

لطالما استخدمت الحواسيب لتصحيح أنواع معينة من المهام الاختبارية، فهناك "آلة تصحيح الاختبارات" التي تحمل اسم "IBM Model 805 Test Scoring Machine" والتي خرجت للنور عام ١٩٣٥م، وتتميز بالقدرة على تصحيح الاختبارات "الموضوعية" بسرعة تفوق سرعة البشر عشر مرات، وبمعدلات عالية من الدقة. (الشكل رقم ٢,٦) ولا يزال هذا المفهوم قيد الاستخدام حتى الآن، ولم يتغير في أصوله إلا فيما يتعلق بصور التقدم في تقنية الحاسوب، علمًا بأن الحجج التي صاحبت استخدام اختبارات الاختيار من متعدد – سواء أكانت معه أم ضده – لم يطلها إلا تغيير ضئيل عما كانت عليه قبل ٨٠ عامًا (Fulcher, 2000). وتكفي الإشارة إلى أنه عندما يتعلق الأمر بتصحيح اختبارات الاختيار من متعدد، تتمثل أوجه الاختلاف بين الإنسان والآلة في عنصري السرعة والدقة، أما لدى تقييم مهام إنتاج النصوص المعقدة، مكتوبة كانت أم مقروءة، فتكون الاختلافات بين البشر والآلات أكثر تعقيدًا وأحرى بالسبر والبحث. ونجد في هذا المقام أن ثلاثة أوجه للمكون المطلوب تقييمه المعني بخصائص طريقة الاختبار ذات صلة وهي: تحديد الجانب اللغوي المطلوب قياسه، ومعايير تقييم الأداء، والإجراء المتبع في تصحيح الأداء ونتائجه.

عند تقييم استجابة مكتوبة أو مقروءة على مهمة اختبارية، يقتضي من البشر إتقان التراكيب في النص، وترجمة المفردات والوحدات الدلالية الأكبر، بالإضافة إلى تحليل جانب كبير من الاستجابة مع تحديدها في إطار براقماتي لغوي لإيضاح معناها، ثم مقارنتها بمجموعة من المعايير الرسمية لتحديد مستوى الجودة، على أن يأتي كل هذا في إطار من القيود التي يفرضها السياق التواصلي قيد النظر. وفي ظل التعقيد الذي يكتنف هذه العملية، يمثل تدريب الأفراد الذين يحتاج إليهم لتقييم القدرات اللغوية بالاختبارات تحديًا أمام جهود البحث والممارسة في هذا الصدد. وللمساعدة في ضمان

الاتساق والتناغم عند تصحيح مستويات الأداء اللغوي المعقدة، لا بد من وضع قواعد دقيقة للتصحيح وتدريب القائمين على التقييم بغية

الشكل رقم (٢,٦). آلة تصحيح الاختبارات (IBM Model 805).

(IBM Corporation, 2002)

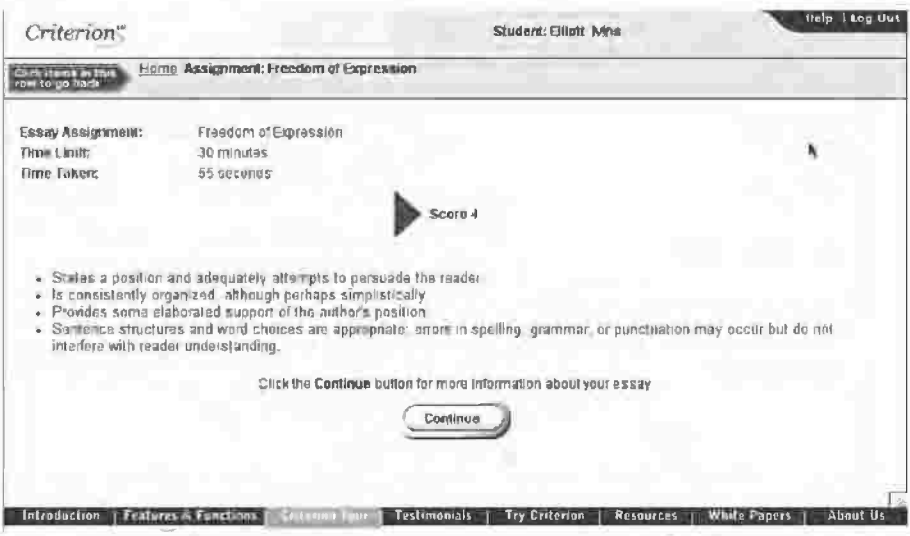
استخدام تلك القواعد على نحو يتسم بالثقة والفاعلية. وكلما كان برنامج الاختبار كبيراً، زاد العدد المطلوب من القائمين على التقييم؛ وبناء عليه زادت تكلفة العملية برمتها. ولم يعد مستغربًا أن تكون الآفاق المرتبطة بقدرة الحاسوب على حساب درجات الاستجابات وتحليلها مصدر جذب مستقبلاً عندما تُطلَب الإجابات القصيرة أو المطولة في اختبارات اللغة سواء أكانت مكتوبة، كما هي الحال في مهام اختبار فهم المقروء (التتمة المنتظمة) أو اختبار المقال، أو الشفهية، كما هي الحال في مهام اختبارات النطق أو محاكاة المقابلات الشفهية. ولتقييم القدرات اللغوية المقروءة أو المكتوبة بالحاسوب، ثمة تقنيات معقدة نسبيًا لا بد من توافرها للتعرف على الكلام وتفسيره وتحليل خطابه السياقي. وقد حدث بعض التقدم في تطبيق تلك الجوانب على التقييم اللغوي، ولذا أضحى بإمكاننا

إمعان النظر في المثالين المذكورين في الفصل السابق، ألا وهما نظام "الكرايتيرين" ونظام "تقييم مهارة التحدث التلقائي "المعروف باسم "Phone Pass".

كرايتيرَيْنْ

تمكن مركز "خدمة الاختبارات التعليمية" الذي يقدم اختبارات "التوفل" وعددًا من الاختبارات المهنية الأكاديمية الأخرى – من تطوير نظام آلي (مؤتمت) يحمل اسم "كرايتيرين" لتقييم الاستجابات المكتوبة المطولة. وهذا النظام يوظف أساليب معالجة اللغات الطبيعية للتفسير والربط التراكيبي للنصوص المدخلة مسندًا بذلك المسميات إلى مكونات النص في إجابة الطالب، ومحددًا علامات البناء الخطابي (النص) ومفردات المحتوى. كما أنه مستخدم في الخدمة المقدمة عبر الإنترنت التي توفر التقييم والتعقيب على الكتابة للطلاب والمعلمين (Educational Testing Service, 2005a). وجدير بالذكر أن هذا النظام قادر على إجراء تقدير شامل لأي مقال استنادًا إلى ميزان مكوَّن من ست نقاط، كما هو موضح في المثال المقدم عبر الإنترنت لتقرير كرايتيرين الشكل رقم (٢,٧). أما التعليق المصاحب للتقدير، فقائم على جوانب التحليل من خلال معالجة اللغات الطبيعية كتلك المستمدة من منتج آخر لاختبار اللغة يحمل اسم "المقدِّر الإلكتروني" (e-rater) (Monagh and Bridgeman, 2005) الذي يتولى تحليل سبع فئات من الظواهر محل التقييم: ومن ذلك؛ آليات القواعد (كالتهجي) والاستخدام (كأخطاء استخدام الأدوات) والأسلوب (ككثرة تكرار الكلمات أو إطالة الجمل إطالة مملة)، وكذا مستوى المفردات المستخدم والتنظيم/التطوير (كتحديد الجمل المناظرة للخلفية، والطرح، والفكرة الأساسية/والفكرة الداعمة والخاتمة) واستخدامات المفردات المنبثقة عن مطالبات محددة. ووفقًا للبحوث التي أجريت على المنتج سالف الذكر، يمكن للمقدر الإلكتروني أيضًا أن يكون بمثابة داعم للشخص القائم بالتقدير، لا أن يكون مجرد أداة آلية للتصحيح قائمة بذاتها (Monaghan and Bridgeman, 2005)، فعلى سبيل

المثال، يمكن أن يقوم الشخص القائم بالتقدير (المُقَيِّم) والمقدر الإلكتروني بتقييم المقال، فإذا كانت النتائج على درجة عالية من التقارب، فإن الدرجة النهائية تتمثل في التقييم البشري وحده، أما إذا اختلف التقديران، فهنا يتدخل مقدر بشري ثان للفصل في هذا التناقض (يتناول الفصل التالي مقارنات بين تقدير الأداء اللغوي البشري والحاسوبي). وقد انتهي كل من "موناغان وبريدجمان" أنه في الحالة الثانية، اقتضت الحاجة اللجوء إلى المقدر البشري الثاني في نحو ١٥٪ من الحالات، وأن استخدام المقدر الإلكتروني كأداة مرجعية على هذا النحو، أسفر عن نتائج جاءت في المتوسط (استنادًا إلى مرجعية من ست نقاط) من درجة معيارية مستقلة للتقييم قدرها ٧٥,٥٪ من الزمن وفي إطار نقطة واحدة بنسبة قدرها ٩٣,٩٪ من الزمن. وهو ما يعني أن تقنية المقدر الإلكتروني، إنما هي تقنية واعدة في المستقبل بصفة عامة من حيث إمكانات التطوير والبحث المتواصلة بغية تعزيز الاستفادة منها على الوجه الأمثل.

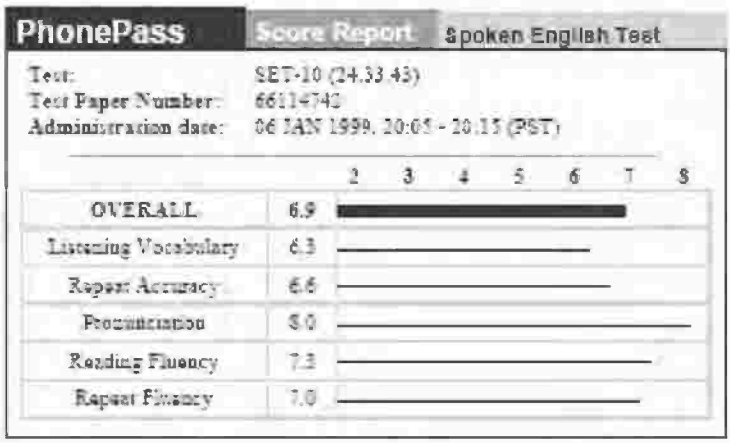


الشكل رقم (٢,٧). مثال على تقرير من نظام كرايتيريَنْ.

(Educational Testing Service, 2005b).

تقييم الاستجابات الهاتفية حاسوبياً

هذا النظام الذي يرمز له بـ "SET-10" (وكان يسمى Phone Pass سابقًا) هو اختبار يصحح حاسوبياً لمتحدثي اللغة الإنجليزية من غير أهلها، وهو من إنتاج شركة "أوردينيت" (Ordinate, Inc.)، ويستفيد من تقنية معالجة اللغات الطبيعية. ويتخلل عمل هذا النظام إعطاء مجموعة من التعليمات المطبوعة للطلاب، ثم التواصل مع الحاسوب عبر الهاتف لتلقي توجيهات شفهية إضافية. وعلى ذلك، يقدم الطلاب استجاباتهم الشفهية عبر الهاتف، ثم تقدر هذه الاستجابات حاسوبياً، علمًا بأن الاختبار يتألف من أنماط متنوعة من المهام التي من بينها قراءة الجمل وتكرارها ونطق الكلمات، وكلها مهام تستلزمها المقابلات الدلالية أو الأسئلة أو باجتماعهما. أما البنود الفعلية، فتؤخذ بصورة عشوائية في كل تطبيق من بين مجموعة هائلة من البنود السابق اختبارها. وبعد ذلك، تعلن النتائج عبر الويب، كما هو موضح في الشكل رقم (٢,٨).


PhonePass Score Report Spoken English Test

Test: SET-10 (24.33.43)
Test Paper Number: 66114742
Administration date: 06 JAN 1999, 20:05 - 20:15 (PST)

		2 3 4 5 6 7 8
OVERALL	6.9	
Listening Vocabulary	6.3	
Repeat Accuracy	6.6	
Pronunciation	8.0	
Reading Fluency	7.3	
Repeat Fluency	7.0	

الشكل رقم (٢,٨). تقرير نتيجة صححت بنظام SET-10.

(Ordinate Corporation, 1999)

يعتمد التقدير بالنظام المذكور (SET-10) على نظام التعرف على الكلام؛ وهو نظام يشتمل على لوغاريثم مستمد من حصيلة كبيرة من اللغة المتحدثة على لسان أهلها، وقد روعي فيها مختلف اللهجات الإقليمية والاجتماعية للغة الإنجليزية. ويتيح النظام ذاته مقارنة مستويات أداء الطلاب بقالب يمثل المفردات المحددة المطلوبة، وكذا الخصائص الصوتية للصوامت والصوائت، بالإضافة إلى صياغة الجمل وتوقيت تقديمها. وقد خضع برنامج التقدير هذا لدراسة مكثفة أسفرت عن ارتباط قدره ٩٤, مع المقدر البشري في التقدير الشامل، وارتباط يتراوح ما بين ٧٩, و٨٩, في النتائج الفرعية المترتبة عليه (Ordinate Corporation, 2002a). وبذلك نخلص إلى أن هذا النظام هو برنامج مستقل للتقدير، يعمل دون تدخل العنصر البشري على الرغم من أنه يخضع لمراقبة دقيقة من إدارة البرنامج.

ثمة طريقة أخرى للاستفادة من إمكانات تقنية الحاسوب للمساعدة في عملية التقدير، تتمثل في توفير أنماط مختلفة من صور المساعدة التي يسهل الوصول إليها للذين يضطلعون بعملية التقدير. ومثال ذلك القائمون على تطوير "اختبار الكفاية الشفهية في اللغة الإنجليزية" (Oral English Proficiency Test) في جامعة "بوردو" (Purdue University) الذين يعملون على نظام حاسوبي للتقدير يقدم البند الاختباري على شاشة تضم خيارًا لتشغيل استجابات معيارية على أي من المهام في ظل انشغال المقدرين (من البشر) بتقدير الاستجابات على الاختبار (Ginther, 2001، تحت عنوان "التواصل الشخصي"). وتجدر الإشارة إلى أن أداء المقدرين يراقب تلقائيًا، أما البنود الخلافية والمقدرون فيخضعان للترميز. ومن المعلوم أن الهدف من وراء ذلك كله أن يسهم النظام في الحفاظ على المعايير التي يحددها المطورون، وبالمثل نجد أن الأداة الشفهية الحاسوبية لإتقان اللغة (COPI) المستحدثة في مركز اللغويات التطبيقية (Center for Applied Linguistics) (Malabonga, Kenyon & Carpenter, 2002) تستخدم نظامًا

حاسوبياً للتصحيح، وهو نظام يتيح للمقدرين الاستماع لاستجابات الطلاب عبر الإنترنت، وجدولة النتائج تلقائيًا استنادًا إلى لوغاريثم محدد مسبقاً، بالإضافة إلى تخزين الاستجابات والتقديرات الخاصة بالطلاب. مما سبق نعلم أن كلا المثالين ينطويان على توجهات مستقبلية واعدة للجمع بين إمكانات المقدرين من البشر وأدوات التقدير الحاسوبية بهدف تقييم الاستجابات الإنشائية (الكتابية).

وعلى ذلك تمثل آفاق التقدير الآلي (المؤتمت) لمهام الاستجابات المقدمة، أحد الجوانب الواعدة والمثيرة في اختبارات اللغة المحوسبة؛ نظرًا لما تنطوي عليه من احتمالات بزيادة الجانب العملي من الاستجابة على المهام الاختبارية اللغوية. وفي ظل هذه الاحتمالات والآفاق البيّنة، تتجلى الحاجة للبحث سعيًا إلى فهم أشمل لقيود معالجة اللغات الطبيعية، لاسيما فيما يتعلق بالنتائج المترتبة عليها في فهم المكونات التي نحاول قياسها. يختلف التصحيح الآلي (المؤتمت) اختلافًا واضحًا عن التصحيح البشري، بيد أن التحدي القائم يتمثل في فهم الظروف التي يكون فيها الأول أكثر إفادة من الثاني لدى تقدير النتاج اللغوي للطلاب.

خاتمة

تناول هذا الفصل السبل التي من خلالها توسع تقنية الحاسوب من الخيارات أمام معدي الاختبارات من أجل تصميم المهام الاختبارية اللغوية، كما قدم العديد من الأمثلة الموضحة لأوجه اختلاف اختبارات اللغة المحوسبة – في جوانب أساسية للطلاب، ومعدي الاختبارات، والمدرسين، والباحثين – عن الأنماط الأخرى لتقديم الاختبارات. وانتهينا إلى أن أوجه الاختلاف التي تقدمها اختبارات اللغة المحوسبة تؤثر في جوانب تتخطى عنصر الفاعلية في اختبار عدد أكبر من الطلاب، وتقييم الاختبارات، والإعلان عن النتائج بسرعة أكبر. علمًا بأن اختبارات اللغة المحوسبة ذات التكلفة الزهيدة تختلف

عن الأنماط الأخرى من اختبارات اللغة في خصائص الطريقة المستخدمة في التعاطي مع المدخلات والاستجابة والتفاعل فيما بينها، إلى جانب الاختلاف في عنصر التقييم. كما وجدنا أن المدخلات السياقية الثرية، وتنوع أساليب الاستجابة، وخاصية التكيف التي يوفرها الحاسوب، والتصحيح الآلي (المؤتمت) الذي صار ممكناً بفضل الحاسوب، تتطلب في مجملها المراعاة الدقيقة لجوانب اللغة التي يريد معدّ الاختبار قياسها، كما تتطلب الأخذ بظواهر الموقف الذي تستخدم فيه اللغة الهدف.

وبالتركيز على الفرص المتاحة من خلال التقنية، لم يتناول هذا الفصل إلا بعض المحذورات التي يجب أن تواكب النظر في اختبارات اللغة المحوسبة، بما في ذلك القضايا الأمنية للاختبارات ذات الاعتبارات والمخاوف العالية، وكذلك مستوى الخبرة الفنية، ومدى التوحيد والتناسق والصيانة اللازمة للأجهزة، إلى جانب احتمال الابتعاد عن عنصر الاتساق الذي يصاحب قدرة الحاسوب على توفير مستويات مختلفة وتوجيهات بلغات مختلفة، مصحوبة بشاشات للمساعدة وأشكال للمدخلات والاستجابات. كما تناول الفصل القيود الحالية على قدرة الحاسوب إزاء محاكاة معالجة اللغات الطبيعية والتفاعل في موقف الاختبار. أما الفصل التالي، فسيتناول هذه القضايا بتوسع أكبر من منظور يسعى إلى الوقوف على كيفية تأثيرها في عنصر الصدق في الاستنتاجات المنبثقة عن اختبارات اللغة المحوسبة واستخداماتها.