الفصل الأول

مسار التقنية
The Technology Thread

من الصعوبة الوقوف على أعداد متعلمي اللغة الثانية الذين تقدموا أو سيتقدمون لاختبارات في اللغة تقدم بالحاسوب، غير أن كثيراً من الاختبارات المصيرية أو غير المصيرية (High-and low-stakes tests) تُقدم الآن من خلال الحاسوب، كما أن أعداد هذه الاختبارات في ازدياد مطّرِد. وهذه الحقيقة القائمة في مجال اختبارات اللغة تتجلى عمليًا في صورة مسار ينتظم فى كل أعمال "سلسلة كمبريدج للتقييم اللغوي". فما من مؤلف من مؤلفي هذه السلسلة، إلا وقد رأى دورًا للتقنية في التقييم[1] اللغوي؛ لاسيما في المستقبل. فعلى سبيل المثال، قام "ريد" (Read, 2000) بتضمين الاختبار المعروف باسم "اختبار المراكز الأوروبية المعتمد على الحاسوب لحصيلة المفردات اللغوية" (Computer- Based Eurocentres Vocabulary Size Test) وكذا اختبار اللغة الإنجليزية كلغة أجنبية في دراسته المسحية لحصيلة المفردات اللغوية وتقييماتها. ويذهب الباحث لما هو أبعد من هذا الطرح المتعلق بالاختبارات المحوسبة، إذ يوضح أن الطرق

---

[1] فرق المترجم في كتاب سابق بين عمليتي التقييم والتقويم، فالتقييم يعني إصدار حكم في ضوء معايير محددة كأن نقيم الطالب بأنه ممتاز، جيد أو ضعيف. أما التقويم فإنه لا يقتصر على إصدار حكم فحسب، بل يتجاوز ذلك إلى اتخاذ إجراءات معينة لتلافي نقاط الضعف لدى الطالب والعمل على تحسين مستواه (المترجم).

والأساليب المعتمدة على الحاسوب تشكل عنصرًا أساسيًا لفهم المفردات المطلوبة للارتقاء بجوانب تقييم حصيلة المفردات اللغوية. وعلى النهج نفسه، يرى "بك" (Buck, 2001) أن تقديم اللغة الشفهية المدعومة بوسائط متعددة – تعتمد على الحاسوب – يعد حيويا" لمستقبل تقييم القدرة على فهم المسموع. أما "ويجل" (Weigle, 2002) فقد طرح تقييم مهارة الكتابة مستقبلاً على بساط البحث، متطرقًا إلى الأساليب القائمة على التقنية لتقييم مهارة الكتابة – كتقييم النص المكتوب بالحاسوب – وآثار التقنية في هذه المهارة. وأما "آلدرسون" (Alderson, 2000) فقد اقترح إعداد اختبار على نطاق واسع عبر الإنترنت، وطرق لاختبار مهارة القراءة عن طريق الحاسوب، بالإضافة إلى إيجاد طرق للقراءة عبر الإنترنت. وأخيراً، بحث "دوجلاس" (Douglas, 2000 : ص ٢٧٥ وما بعدها) في موضوع "مثالب التقنية" في عمل له بعنوان "تقييم اللغات لأغراض خاصة".

وبأخذ كل ما أوردناه آنفاً في الحسبان، نرى أن ما تحقق حتى الآن في مسار التقنية ينبئ بأننا مقبلون على تغيير مهم في صميم عملية التقييم اللغوي: ألا وهو الاستعانة بالتقنية على نحو واسع. وسنتناول في هذا الكتاب أهمية التطورات التي تحدثها الاستخدامات الحديثة للتقنية فيما يتصل بالتقييم اللغوي، كما سنستكشف صور التغيير في المعارف المهنية المطلوبة لاستخدام تلك التقنية. وسنمضي في فصول الكتاب مستخدمين لفظتي "اختبار" و"تقييم" بصورة تبادلية في إطار بحثنا لمجموعة كاملة من الاستخدامات المصيرية وغير المصيرية للتقييمات المعتمدة على التقنية فيما يتعلق ببناء مهام الاختبار وتصحيح الأداء الذي يقوم به الطلاب. بيد أننا لم نُدخِل في ذلك العديد من استخدامات الحاسوب الأخرى لأغراض معالجة البيانات، والتحليل الإحصائي مما نشير إليه باسم الاختبارات التقليدية أو غير الحاسوبية (Davidson 1996, Bachman 2004 على التوالي للاطلاع على بعض الآراء المتعلقة بهذا الموضوع). وفي المجالات ذات الصلة

بعلم اللغة التطبيقي كدراسة استخدام اللغة وبحوث اكتساب اللغة الثانية وتدريسها، نجد أن للتقنية آثارًا ملموسة في المعارف والممارسات المهنية في تلك المجالات التي كشفت فيها البحوث والتطبيقات العملية عن توسع التقنية وإسهامها في تغيير الحاجات المفاهيمية والعملية لدى مستخدميها. كما أن المحاولات الرامية إلى تلبية المطالب الجديدة عادة ما تتيح سبر أفهام المستخدمين لعملهم في مجال علوم اللغة التطبيقية.

وعلى النحو ذاته، فقد ازدادت الجهود في استكشاف دور التقنية في تقييم اللغة لأغراض الاختبارات إلى الحد الذي نعلم به اليوم أن دارسي اللغة الثانية سيخضعون عاجلاً أم آجلاً لاختبار اللغة بواسطة الحاسوب في أي مكان من العالم. ومن المعلوم أن اختبار "التوفل" (اختبار اللغة الإنجليزية بوصفها لغة أجنبية "TOEFL") — الذي يعد من أفضل برامج اختبارات اللغة الثانية وأوسعها انتشارًا في العالم — يقدم عبر الحاسوب في الكثير من الدول، ويتقدم له مئات الآلاف من الطلاب سنويًا (خدمة الاختبارات التعليمية، برنامج التوفل: http://www.ets.org/toefl/). وعلى النهج نفسه، تقدم الكثير من فصول تعليم اللغة، وكذلك المواقع التي تعنى بمواد تعليم اللغة على شبكة الإنترنت، تقييمات هي بمثابة اختبارات تشخيصية أو تحصيلية، ومن ذلك موقع Longman English Interactive (Rost, 2003). وبناء على ذلك، فإن التقييم المحوسب الذي أشير إليه في الكتب الأخرى من هذه السلسلة، فضلاً عن العدد المتزايد من الاختبارات والبرامج التعليمية التي تتيح إمكانية التقييم عبر الإنترنت، تشير في مجملها إلى أهمية التقنية في التقييم اللغوي في مستقبلاً. ومن ثمّ، فإننا بصدد التوسع في استقراء هذا الكتاب من خلال مناقشة ما يمكن أن يُضْفيه استخدام الحاسوب لمجال التقييم اللغوي، علمًا بأن أكثر الفئات تأثرًا بهذه التغييرات إنما هم المتقدمون لهذه الاختبارات؛ وذلك لأنهم المستخدمون لتلك التقنية في نهاية المطاف. أما شريحة القراء التي يستهدفها هذا الكتاب، فهم الخبراء الساعون لمساعدة المتعلمين، ولذا سنستهل

هذا الفصل بتوضيح بعض الملابسات التي تكتنف اختبار اللغة المحوسب للمدرسين، ومعدي الاختبارات، والباحثين في مجال اختبارات اللغة.

مدرسو اللغات

يحتاج مدرسو اللغات إلى فهم حقيقي سليم لقضية التقييم اللغوي؛ ذلك أنه منوط بهم مساعدة متعلمي اللغة لإكسابهم أساليب وإستراتيجيات للتقييم الذاتي، وكذلك اختبارهم في الصفوف الدراسية، واختيار اختبارات برامج اللغة أو إعدادها، علاوة على تهيئة المتعلمين للاختبارات التي تعقد خارج الصفوف الدراسية وبرامج اللغة. ومما لاشك فيه، أن كثيرًا من المدرسين يشعرون بشيء من القلق، بل والغضب أحيانًا؛ لاعتقادهم بأن إجراء اختبارات من هذا النوع ، وخاصة المصيرية منها، عبر الإنترنت قد يعطل ملكات المتعلمين ويثنيهم عن إظهار كامل قدراتهم. أما الجوانب التي تتعلق بمسألة العدالة والإنصاف بين جميع الطلاب، فما هي إلا بعض المخاوف التي تكتنف استخدامات التقنية في الاختبارات. وثمة جوانب أخرى من بينها المعارف اللازمة لاختيار الاختبارات المحوسبة وإعدادها واستخدامها. وفي الوقت ذاته، بمقدور المدرسين والمتعلمين الاستفادة من هذا النوع من الاختبارات من خلال الوصول إلى الاختبارات التشخيصية واختبارات تحديد المستوى التي يمكن أن تكون متاحة على شبكة الإنترنت، ومن ثمَّ إمكانية تحليل استجابات المتعلمين والتعقيب عليها والاحتفاظ بسجلاتهم، وهو ما يعني مزيدًا من الإمكانات والمعطيات التي لم تكن متاحة من خلال أساليب التقييم التقليدية.

اختيار الاختبارات

هناك احتمال بأن تتاح الفرصة للمدرسين بأن ينتقوا من بين مجموعة واسعة ومتنوعة من الاختبارات المحوسبة، وتقتضي الحاجة هنا الإلمام بالكيفية المثلى لتقييم

مثل تلك الاختبارات. وثمة سؤال يطرح نفسه هنا: هل الإرشادات التوجيهية المستمدة من القياس التربوي لتحليل ثبات الاختبار، وصدقه، وسهولة تطبيقه، وأصالته تشمل -على سبيل المثال- كل الاعتبارات المتعلقة بالتقييم المعتمد على الحاسوب؟ ففي الحالات التي شهدت تقييم المواد اللغوية حاسوبياً (Susser, 2001) اقترح استخدام قوائم التحقق للاختبارات المعتمدة على الحاسوب (Noijons, 1994) شاملةً عناصر قد يجدها المرء مدرجة على أي قائمة من قوائم التحقيق للوقوف على جودة أي اختبار (كعنصر وضوح المعلومات مثلاً) مصحوبة بتعديلات تمس الجانب التقني (كالمعلومات المتاحة بشأن خيارات المساعدة والتعليمات مثلاً). أما الخيارات الأخرى؛ فهي فريدة في طبيعتها فيما يتعلق بالأجواء المكانية والزمانية للاختبار المحوسب (كسرِّية بيانات الاستجابة على الاختبار فور إتمامه مثلاً). وهذه القوائم مستقاة -في المقام الأول- من حقل القياس التربوي (مثال ذلك Green, 1988)، ولذا يتوقع أن تشكل أساسًا متينًا من حيث النتائج المرجوة منها. غير أنه ينبغي لنا أيضًا أن نتساءل بشأن مدى استيعابها لكل الأمور المتعلقة بالتقييم اللغوي. ونضرب لذلك مثالاً مفاده أن الاختبارات في المجالات الأخرى، نادرًا ما تضم أيُّ اختبار شفوي، الأمر الذي ينطوي عليه احتمال ضآلة الاهتمام والتحليل الموجهين إلى مهارتي التحدث والاستماع عبر الحاسوب في ظل الأطر سالفة الذكر. وسنتناول موضوع تقييم اختبارات اللغة المحوسبة في الفصل الخامس من هذا الكتاب.

التقييم الصفي

تزداد الفرص أمام المدرسين باطراد للاستفادة من اختبارات اللغة المحوسبة الموجودة بوصفها جزءاً من الدورات اللغوية المتوفرة على شبكة الإنترنت، أو الاستفادة من برامج التوليف (authoring software) التي تتيح لهم إعداد اختباراتهم بأنفسهم.

وهذه السبل الرامية إلى التقييم في الصفوف الدراسية، تقدم إمكانات طيبة لتقييم معدلات التعلم لدى الطالب بانتظام، مصحوبة بإمكانية التعقيب على أدائه بصورة مفصلة. ولهذه الإمكانات أهميتها، حيث تعد أحد عوامل التحول نحو الاختبارات المحوسبة منذ الأيام الأولى لاستخدام التقنية في تعليم اللغة (Otto, 1989).

ويؤكد ذلك شواهد عدة، لعل أسبقها ظهوراً المنهج الفرنسي الموجود في نظام المنهج المبرمج لعمليات التدريس المعروف باسم بلاتو (PLATO) في جامعة "إيلينوي"، وهو النظام الذي احتفظ بسجلات لأداء المتعلمين خلال كل جلسة من جلساتهم طوال مدة الدورة التعليمية، وعلى مدى فصل دراسي كامل، ويتيح هذا النظام أيضا للمتعلمين معلومات موجزة بشأن معدلات أدائهم عند طلبها (Marty, 1981)، وفي هذا الإطار، قدَّم الباحث مارتي مثالاً يحمل اسم "تحليل الأداء العام" (General Performance Analysis) حيث يتيح للمتعلم طلب نتائج تحليل أدائه في أي وقت خلال الفصل الدراسي. ويعرف الطالب من هذا التحليل -على سبيل المثال- أنه قد أتم دراسة ٢٩٨ بابًا من أبواب قواعد اللغة، وأنه حاز في المعدل الكلي على نسبة ٧٧٪ من كل تلك الأبواب. ويمكن للمتعلم متى شاء، الحصول على تحليل بتفاصيل أوفى بمجرد طلبه الاطلاع على الأبواب التي حصل فيها على أقل من ٤٠٪. ويوضح الشكل رقم (١,١) نوع التعقيب أو التغذية الراجعة التي ظهرت على الشاشة عند رفع هذا الطلب إلى النظام، علمًا بأن العمود الأول يشير إلى رمز الباب المدروس في مهارة القواعد، في حين يشير العمود الثاني إلى الإجابات الصحيحة بالنسبة المئوية، أما العمود الثالث فيشير إلى عدد البنود التي استكملها المتعلم في الباب المحدد. وعلاوة على رمز الباب المدروس، حصل المتعلمون على وصف لكل مبحث في كل باب يقفون عليه باتباعهم التعليمات التي توجههم للتعامل مع النظام.

Assez with modifier	12	30%	1
De-verb + partitive	13	30%	21
Verb + de + infinitive	19	20%	37
Ne pas not split with infinitive	14	10%	42

اضغط DATA (بيانات) لإدخال رمز مختلف

اضغط SHIFT-LAB لاستعراض بند في القواعد

الشكل رقم (١,١). تحليل أخطاء المتعلمين من مواد تعلم الفرنسية.

(مارتي، 1981: ص ٣٩)

وُضعت هذه التقييمات التشخيصية على مدار ٢٠ سنة، في ظل بيئة تعاونية بين البحوث المخصصة لتعليم الفرنسية وتعلمها، وسبل التطوير المقترحة في هذا الصدد. غير أن التعقيد المتأصل في التقييم التشخيصي بالحاسوب يقتضي الالتزام بنهج بحثي مستدام لا مجرد مشروع واحد مؤقت، كما هي الحال مع مشروع DIALANG (Alderson, 2000). من جانب آخر، شرع الناشرون التجاريون – ممن لديهم القدرة على تطوير المواد المعقدة على الإنترنت – في الاقتراب من بعض هذه الأفكار بحثًا عن عناصر التقييم التشخيصي أو المنجزات المصممة للتوافق مع متطلبات الدورات الدراسية. ولا تفوتنا هنا الإشارة إلى أن دورات اللغة الإنجليزية المقدمة عبر الإنترنت: كدورات "ماركت ليدر" (Market Leader) (Longman, 2002) تحتوي على جزء مختص بالتقييم التكاملي على مدار الدورات، بغية تقديم معلومات للمتعلمين والمدرسين على حد سواء قبل الاختبارات وبعدها. وهذه الاختبارات تعد وتصمم وفق المبادئ المعروفة بالاختبارات محكية المرجع، لكن المثال المأخوذ من دورات اللغة الفرنسية، يؤكد أن تلك المبادئ الأساسية قد تستنفد عن آخرها بصورة مختلفة لدى إعداد الاختبارات عبر الإنترنت.

وحيثما أتيحت فرصة تعليم اللغة على الإنترنت، يصبح من نافلة القول أن يفكر المدرسون في الأخذ أيضًا بجوانب التقييم عبر الإنترنت على الأقل، لكن بعض الاختبارات الفردية قد تأتي في أفضل صورها عبر الحاسوب، عندما يتطلب إعدادها معلومات تشخيصية مفصلة. وبالمثال يتضح المقال، فمنذ عدة سنوات طالعنا كل من "مولهولت وبريزلر" (Molholt and Presler (1986 باقتراحهما إزاء إمكانية استخدام تحليل النطق الذي صممناه؛ للوقوف على الجوانب المحددة للنطق التي تستلزم توافر تعليمات بعينها. أما بحث "كانيل" (Canale)(1986) فقد احتوى على مقترحات ترى الأخذ بأنظمة ذكية للتعليم يكون بمقدورها جمع معلومات تشخيصية عن المتعلمين لدى مباشرة التعلم عبر الإنترنت. علمًا بأنه قد وصف عدداً من هذه الأنظمة لتعلم اللغة، لكن هذا البحث قد أكد على الإمكانات التعليمية لهذه الأنظمة دون سبر أغوارها سبرًا وافيًا على صعيد التقييم (على سبيل المثال Holland, Kaplan & Sams,1994). وعلى الرغم مما سبق، فإن استكشاف المعلومات المفصلة المستمدة من التقييمات التشخيصية تنطوي على تحديات مثيرة في مجال التقييم اللغوي باعتباره مجالاً قائمًا بذاته، وفي هذا يوضح "كلارك" (Clark,1989) أن الاختبارات التشخيصية تُصمَّم وفقًا لمواصفات مختلفة عن مواصفات إعداد اختبارات الكفاية اللغوية ، وهذا الأخير إنما يُطلب له درجة واحدة لا أكثر. بيد أن جانبًا كبيرًا من المعارف النظرية والعملية المتوفرة بشأن وضع التقييمات وتفريغ نتائجها، قد وُضعت في المقام الأول كاختبارات لتقييم الكفاية اللغوية تاركةً جوانب التشخيص مجالاً بكراً غير مطروق إلى حد ما. ومع تزايد أعداد الأشخاص القادرين على إعداد التقييمات التشخيصية حاسوبياً واستخدامها، فإنه من المرجح تناول هذه الجوانب بفهم أوفى وأرحب بمرور الوقت (انظر: (Alderson, 2005 للاطلاع على المناقشة التي تمت في هذا الخصوص).

إعداد الاختبارات

عادة ما يضع المدرسون اختبارات التقييم الصفي بأنفسهم ؛ بما يعكس النقاط المهمة التي دُرّست للطلاب، وعلى ذلك تتوافر مجموعة من الخيارات لدى المدرسين الراغبين في إعداد الاختبارات على شبكة الإنترنت بأنفسهم. ولعل أنجع هذه الخيارات يتمثل في برامج إدارة الدورات الدراسية (course management software) التي تتيح للمدرس بناء وحدات تقييميه تشتمل على اختبارات قصيرة ؛ أي وضع أسئلة بعينها تقدم في سياق اختبار قصير مصحوب بأداة للتقييم، وإبلاغ الطلاب والمدرسين بالنتائج. ومثل هذه البرامج المؤلفة مفيدة للغاية فيما يتصل بتزويد المدرسين بإمكانية الوصول إلى الخطوات الفعلية لإعداد الاختبار ؛ فهي لا تتطلب سوى قدر يسير من التدريب. لكن عنصر الكفاءة – كما سنرى في الفصل الرابع – عادة ما يتوفر على حساب جوانب أخرى محددة لها أهميتها أيضًا مثل: تنوع بنود الاختبار، وتحليل الاستجابة ذات الحساسية اللغوية. غير أننا نجد في تلك البرامج المؤلفة ذات الأغراض العامة – إجمالاً – فرصة للمدرسين للوصول إلى الخطوات الفعلية لإعداد الاختبارات وإلى بعض إمكانيات اختبارات اللغة المحوسبة.

ونتيجة لذلك، يمكن أن تتضافر جهود المدرسين –بالعمل سوياً – لإعداد اختبارات تتناسب مع البرنامج التعليمي المعمول به لديهم، فعلى سبيل المثال، استحدث معهد تعليم اللغة الإنجليزية (ELI) بجامعة "سوري" بالمملكة المتحدة، حذف عدد من أنشطة التقييم الذاتي المخصصة لجبر القصور في الدورات التعليمية التي يقدمها. واشتملت تلك الأنشطة على اختبارات قصيرة تقدم للمتعلمين تغذية راجعة فورية بما يتيح لهم تقييم مدى تقدمهم في العملية التعليمية، كما هو موضح في الشكل رقم (١,٢) الذي يقدم مثالاً لاختبار قصير في كتابة الأطروحات العلمية. ومجمل القول إن المدرسين

والطلاب –على السواء– قد يستفيدون من إعداد مثل هذه الاختبارات القصيرة على الإنترنت واستخدامها ؛ نظرًا لعدم تطلبها أدوات أو مهارات معقدة لكتابتها.

الأطروحة ١

شكرًا لتقدمك لاختبار التقييم الذاتي عن الوحدة (١) لكتابة الأطروحة

القسم الأول: الإعداد

في مرحلة الإعداد لأطروحتك العملية، وقبل الشروع فعلاً في بحثك، لا بد من مراعاة عدد من الأنشطة بالغة الأهمية فور تحديد موضوعك. اختر مما يلي أهم خمسة أنشطة:

- تحديد الأهداف كان صحيحًا
- الإجابة الصحيحة كانت كتابة عرض مبدئي لمضمون الأطروحة
- صياغة العنوان تساعد في توضيح الفكرة في بادئ الأمر، حتى وإن قررت العدول عن ذلك لاحقًا.

ثمة حاجة بك إلى تحديد أهداف بأسرع وقت ممكن لتتأكد أن بحثك يمضي باتجاه محدد. وهذا ييسر لك أيضًا اختيار المادة التي ستقرؤها! العرض المبدئي يسهم أيضًا في إبراز القضايا وتوضيحها. محور الأطروحة أمر مهم أيضًا: إذ يجب ألاّ يكون فضفاضًا ولا محدودًا للغاية. وأخيرًا، من المهم دائمًا وضع جدول زمني بمواعيد محددة لإنهاء العمل.

الشكل رقم (١,٢). التغذية الراجعة لاختبار ذاتي قصير في كتابة الأطروحات العلمية في معهد ELI بجامعة "سوري" (http://www.surrey.ac.uk/ELI/sa/thesisl.html)

معدو الاختبارات

يتعامل المحترفون في إعداد الاختبارات المحوسبة، مع قدر أكبر وأوسع من الخيارات مقارنة بما هو متاح عادة عند إعداد الاختبارات التقليدية، بما في ذلك خيارات تقديم الاختبارات التي تتيح مجالات أرحب لتقييم اللغة. وتشمل الطرق الحديثة في هذا الصدد: الاختبار الحاسوبي المتكيف، واستخدام الوسائط المتعددة لطرح المدخلات اللغوية والمرئية للمتعلمين، وكذا التحليل الآلي للاستجابات. وتثير هذه الطرق الحديثة أسئلة لدى معدي الاختبارات بشأن النقاط الخاضعة للتقييم في اختبارات اللغة الجديدة.

الاختبار الحاسوبي المتكيف

يربط كثير من محترفي إعداد الاختبارات بين الحاسوبيات المخصصة لعرض الاختبارات وبين إعداد مجموعات كبيرة من البنود المستخدمة في الاختبارات الحاسوبية المتكيفة (تعرف اختصارًا باسم CATs). يتولى الاختبار الحاسوبي المتكيف اختيار البنود ومن ثم تقديمها بالتتالي استنادًا إلى استجابة الممتحن لكل بند من البنود؛ فإذا أجاب الطالب عن السؤال الأول إجابة صحيحة، يختار الحاسوب بنداً أصعب من مجموعة البنود ليُقدمه على هيئة سؤال ثانٍ، وإذا أجاب عن السؤال الثاني إجابة صحيحة أيضًا، فسيُعرض عليه سؤال أصعب وهكذا دواليك. أما إذا أخفق الطالب في الإجابة عن سؤال ما، فسيقدم له النظام سؤالاً أسهل من سابقه وهكذا. ومن مزايا الاختبار الحاسوبي المتكيف، أنه يتعرف على مستوى الطالب من خلال مراقبة صعوبة البنود التي تظهر له ويجيب عنها صوابًا أو خطأً، ومن ثم لا يختار له إلا تلك البنود التي تناسب قدراته. وعندما يقدم البرنامج بنوداً كافية للوقوف على قدرة الطالب، استنادًا إلى مستوى محدد مسبقًا عن طريق الثبات، ينتهي الاختبار عندئذٍ، ويتبعه الإعلان عن النتيجة. وبذلك نجد أن هذا النوع من الاختبارات يتميز بالكفاءة؛ نظرًا لتقديم بنود أقرب إلى مستوى الطلاب وقدراتهم، وهو ما يعني تجنب البنود ذات الصعوبة أو السهولة المفرطة التي قد لا تقدم صورة وافية وواقعية عن مستوى الطلاب.

تعرف المشتغلون بإعداد الاختبارات على مزايا الاختبارات الحاسوبية المتكيفة منذ ما يقرب من ٢٠ عامًا. وفي هذا الصدد ذكر "تنج" (Tung, 1986) المزايا التالية: إن هذه الاختبارات تتطلب عددًا أقل من البنود مقارنة بنظيراتها الورقية، كما تتيح تجنب تحدي الطلاب ببنود وأسئلة تفوق إمكاناتهم وقدراتهم؛ ذلك أنها تنتخب لها البنود وفقًا لمستوى مناسب من الصعوبة والسهولة، كما توفر قدرًا أكبر من السرية من خلال الاختيار من بين مجموعة كبيرة من البنود لإعداد الاختبارات الفردية. لقد أصبحت الاختبارات الحاسوبية

المتكيفة ممكنة بفضل التطورات التي شهدتها نظرية القياس المعروفة باسم "نظرية الاستجابة البندية" (Hambleton, Swaniinathan & Rogers, 1991 Lord, 1980) (Item Response Theory) كوسيلة للحصول على البيانات الإحصائية الوافية بشأن بنود الاختبار، وكذا بفضل التطورات في برامج الحاسوب المخصصة لحساب درجات البنود والتحكم في أسلوب تكيفي لاختيار البنود وتقديمها ومن ثم تقييمها (Green, Bock, Humphreys, Linn & Reckase, 1984; Wainer, Dorans, Flaugher, Green, Mislevy, Steinberg & Thissen, 1990; Brown, 1997).انظر أيضًا "باخمان" (Bachman, 2004 – الفصل الثالث) للإطلاع على تصدير مفاهيمي لنظرية الاستجابة للبند.

ومن الأمثلة الأخرى في هذا الإطار تلك الاختبارات التي صممت في مطلع ثمانينيات القرن المنصرم في جامعة "بريجهام يونج" وأعدها كل من "لارسون ومادسون" (Larsonand Madsen, 1985)، كما أعدت اختبارات أخرى لقياس القدرة اللغوية بالاختبارات الحاسوبية المتكيفة على مدار العقد الأخير من القرن ذاته (انظر مثلاً Kaya-Cartoi:, Carton & Dandonoli, 1991; Burston & Monville-Burston, 1895; Crown & Iwashita, 1996; Young, Shermis, Brutten & Perkins, 1996). وقد أثمرت هذه المشاريع العلمية عن إثارة قضايا مهمة بشأن أسلوب قياس القدرة اللغوية، والحاجة إلى بنود مستقلة، وكيفية اختيارها للمختبرين من خلال تكييفها عن طريق الحاسوب. من جانب آخر، جمعت دراسة "تشالو – ديفيل" (Chalhoub-Deville) – كما ظهرت في كتابه المحرر في القرن الماضي – بين اتجاهات نظرية وأخرى عملية لبحث الاختبارات الحاسوبية المتكيفة قرابة ١٢ حالة. وقد أسفرت الأبحاث النظرية عن التأكيد على تعدد الأبعاد في تركيبة هذه الحالات، في حين كشفت الأوصاف الخاصة بإعداد الاختبارات عن نتائج أحادية الأبعاد؛ لاسيما فيما يتعلق بتحديد المستوى (انظر مثلاً: Dunkel, 1999 وLaurier, 1999). وتوحي النتائج المنبثقة عن هذا العمل بأن عملية التكييف بالحاسوب قابلة للاستخدام بهدف إعداد اختبارات فعالة لقياس القدرة اللغوية من أجل الوقوف

على مستوى القدرات في هذا الجانب وما يتصل به من مهارات: كالقراءة وفهم النصوص، لكن معظم المهتمين يتفقون في الوقت ذاته على أن مثل هذه الاختبارات عاجزة عن استغلال ميزة "تعدد القدرات والإمكانات" التي يوفرها الحاسوب.

غير أن فكرة التكييف بالحاسوب تظل تتوسع وتخضع لمزيد من السبر والاستكشاف، ويمكن حاليًا الإشارة إلى أي صورة من صور التفريع أو خيارات المسارات البديلة التي تُنتَخب للطلاب بغية التعرض لها في إطار برنامج قائم على مستوى استجاباتهم. ومثال ذلك، أن الاختبارات قد تتوسع في المستقبل استنادًا إلى أنموذج حالي – ولنذكر "تقييم مؤسسة لونجمان للغة الإنجليزية" (Longman English Assessment) – الذي يتفرع إلى تقييم اكتساب اللغة للأغراض العامة أو لأغراض خاصة كالأعمال؛ وذلك استنادًا إلى استجابة الطالب لاستبيان يحدد رغباته في بداية الاختبار. وفي هذه الحال، يتم تكييف وتهيئة محتوى مدخلات اللغة، ليتناسب مع مجالات اهتمام الطلاب بالقدر المناسب. أما في الحالات الأخرى، فقد يجري تكييف محتوى الاختبارات استنادًا إلى مستوى أداء الطالب في الأقسام السابقة من الاختبار. وخلاصة القول، إن معدي الاختبارات، قد شرعوا لتوِهمْ في سبر غور الأساليب التي تعينهم على تهيئة الاختبار وتكييفه ليتناسب مع الطالب، وهذا مجال تمثل فيه التقنية تحديًا أمام معدي الاختبارات لدى شروعهم في إعداد اختبارات تتناسب واحتياجات المتعلمين واهتماماتهم.

مهام الوسائط المتعددة

ثمة خيار آخر محتمل ينطوي على قوة في نتائجه، ويتمثل ذلك في أن الحواسيب توفر لمعدي الاختبارات مدخلات متعددة النماذج بصورها ومحتواها على شاكلة: وسائط فيديو، أو وسائط نصية، أو صوتية، أو رسومات ملونة؛ وهذا من شأنه أن يعزز من أصالة المدخلات والاستجابات. وفي هذا الصدد، نعلم أن معدي الاختبارات معنيون بتعزيز جانبين من جوانب الأصالة، أولاهما: الأصالة الموقفية (Situational Authenticity)

التي تحدد معنى الأصالة من حيث معطيات السياق لتشمل: أجواء الاختبار، والمشاركين، والمحتوى، والأسلوب، والتخصص. أما الجانب الثاني: فهو الأصالة التفاعلية (Interactional Authenticity) التي تحدد معنى الأصالة من حيث التفاعل بين المعارف اللغوية لدى الطالب ومهمة التواصل محل النظر (Bachman, 1991). ومن المعلوم أن الوسائط المتعددة تُعين أحيانًا على إيجاد هذين الجانبين بصورتيهما البعيدة عن أجواء الاختبار في حال الخضوع له، ومثال ذلك اختبار تحديد المستوى في "نظام التقييم اللغوي عبر الإنترنت" (WebLAS) بجامعة كاليفورنيا في لوس أنجلوس، وهو الاختبار الذي أعد لتوفير معلومات بشأن المستوى، ومدى التقدم، وتشخيص القدرات، والوقوف على إنجازات برامج تعليم اللغات الأجنبية واللغة الثانية بالجامعة، إذ احتوى الاختبار على استخدام الفيديو لتقديم محتوى المحاضرات لأغراض تتعلق بفهم النصوص. وقد تمثل الغرض من استخدام الفيديو في تعزيز جانب "الأصالة الموقفية" في الاختبار من خلال إيجاد ملامح السياق الأكاديمي كالصف الدراسي، والسبورة، وشرائح العرض التقديمي المقدمة ببرنامج العروض التقديمية PowerPoint". ويمكن للمرء أن يتصور مواقف أخرى (غير اختبارية) كمتابعة رحلة مع مرشد، أو الحجز في فندق، أو المشاركة في اجتماع عمل حيث سيضيف عنصر الفيديو مقومات أخرى للسياق غير الاختباري الذي يُفترض أن يعرضه الاختبار. وجدير بالذكر أن برامج الوسائط المتعددة لتعلم اللغات تضم مثل هذه المواقف؛ لتقدم بذلك أمثلة جيدة على الإمكانات المتاحة أمام معدي الاختبارات.

التحليل الآلي للاستجابات

تنطوي الاختبارات التي تطالب الطالب بنتاج لغوي على زيادة عنصر "الأصالة في التفاعل" أكثر مما عليه الحال في الاختبارات التي تتطلب استجابات مختارة؛ لأن النوع الأول عادة ما يتطلب قدرًا أوسع وأشمل من المعارف اللغوية والمعلوماتية،

فضلاً عن ضرورة استخدام بعض الكفايات الإستراتيجية (Strategic Competence) على نحو أكثر تعقيداً. وفي هذا الصدد، بحث بعض معدي اختبارات اللغة قضية استخدام تقنيات معالجة اللغات الطبيعية، لاستحداث طرق لرصد درجات النتاج اللغوي للطلاب، ومن ذلك مثلاً "تقييم مهارة التحدث آلياً" المعروف باسم "PhonePass" (Ordinate Corporation, 2002b) الذي يسجل مدى دقة الكلمات المكررة، والنطق، والطلاقة في القراءة والتكرار استنادًا إلى نظام حاسوبي للتعرف على الكلام وهو يحتوي على لوغاريثم مستمد من حصيلة كبيرة من اللغة المنطوقة على لسان أهلها، وقد روعي فيها مختلف اللهجات الإقليمية والاجتماعية للغة الإنجليزية. من جانب آخر، استطاع مركز "خدمة الاختبارات التعليمية" الذي يقدم اختبارات "التوفل" – إلى جانب عدد من الاختبارات الأكاديمية والمهنية الأخرى – استحداث نظام آلي (مؤتمت) أُطلِق عليه اسم كرايْتيرْيَنْ؛ أي المعيار (Criterion, 2005a) لتقدير الاستجابات الكتابية الموسعة، استنادًا إلى تقنية معالجة اللغات الطبيعية التي تميز كل مدخل من المدخلات، وتحدد المعلومات البنيوية الخاصة بالخطاب من بعض وحدات النص المنتقاة، علاوة على تحليل مفردات الموضوع للخروج بتقويم شامل لأي مقال بناء على مقياس مؤلف من ست نقاط.

طرائق اختبارية جديدة، مكونات جديدة

في أول مجموعة من الأوراق البحثية التي تناولت اختبارات اللغة المحوسبة، أوضح "كانالي" (Canale, 1986) أن استخدام الحاسوب يحمل في طياته وعوداً مستقبلية مؤكدة بتوفير سبل أفضل لقياس مختلف المكونات (القدرات) اللغوية مقارنة بما كان ممكنًا من خلال طرق الاختبارات التقليدية. ومع ذلك، لوحظ اتجاه جهود البحث والتطوير إلى التركيز على أهداف بعينها تتمثل في زيادة معدلات الكفاية والأصالة في الاختبارات. غير أن القليل من البحوث – حتى اللحظة الراهنة – قد تطرق إلى الأسئلة

الشائكة بشأن كيفية استخدام الحاسوب لتقييم القدرات أو المكونات اللغوية التي تختلف عن تلك التي تقيمها الطرق التقليدية حاليًا. وقد بحث "آلدرسون" هذه المسائل موضحًا الإمكانات الحاسوبية ذات الصلة باستكشاف مختلف الجوانب وصولاً إلى آلية ابتكارية للتعامل مع اختبارات اللغة المحوسبة وهي:

١- قدرة الحاسوب على قياس الوقت؛ إذ يمكن قياس الوقت الذي يستغرقه المتعلم لإكمال مهمة معينة، أو الوقت المستغرق في الأجزاء المختلفة من تلك المهمة، وكذا التحكم فيه وتسجيله حاسوبياً.

٢- قدرة الحاسوب على تسجيل معلومات حول اتجاهات الطالب من خلال الاختبار.

٣- قدرة الحاسوب على تقديم معلومات بأساليب مختلفة.

٤- قدرة الحاسوب على توفير الوصول السريع والسهل إلى مجموعة متنوعة من أنواع المعلومات.

٥- إمكانية ربط الحاسوب بأجهزة أخرى، مما يتيح أنواعًا مختلفة من المدخلات وطرائق التقديم.

٦- قدرة الحاسوب على تشجيع المتعلم لاستخدام إستراتيجياته للتقويم. وعلى وجه الخصوص، يمكن للمعلومات التي يجمعها الحاسوب ويقدمها بشأن الأداء في الاختبار أن تخلق لدى المتعلم شعورًا بأهمية آرائه.

٧- قدرة الحاسوب على استخدام قواعد اللغة: أ) فللحاسب القدرة بمستوى ضئيل نسبيًا على تحري الأخطاء الإملائية في النص المقدم من المتعلم؛ ب) يمكن استخدام أدوات تحليل صحة التراكيب (Parsers) بإمكاناتها المتفاوتة؛ ليس للتحقق من صحة التراكيب في نص المتعلم فحسب، بل ولتقديم اختبارات تواصلية أيضًا (Alderson,1990، ص٣٩: ٤٣).

تزامنت جهود "آلدرسون" في إنجلترا - لاستكشاف استخدامات اختبارات اللغة المحوسبة وإمكانياتها وآفاقها - مع جهود "كوربل" في استراليا الرامية إلى الأهداف نفسها، فقد عدَّد الأخير (Corbel, 1993) الإمكانات التي هي بحاجة إلى مزيد من الاستكشاف والتقصي، وصاغها في شكل أسئلة تمثل نقطة انطلاق لجهود بحثية متكاملة، وجاءت على النحو التالي:

١- هل يمكن تلبية الحاجة إلى التنوع في الاختبارات من خلال استخدام نماذج محاكاة حاسوبية، تتفرع وفقًا لمعطيات السياق والغرض؟

٢- هل يمكن تفعيل مفهوم "الاختبارات التواصلية القائمة على المهام" (Communicative Task-Based Tests)؛ على نحو أكثر كفاءة من خلال الحاسوب؟

٣- هل يمكن عند استخدام بنوك سجلات التعقيبات (Comments) وملخصاتها، توفير أسلوب ما يكون قادراً على استيعاب الأبعاد الكثيرة لإتقان اللغة؟ وهل يمكن الحصول على معلومات ذات قيمة تشخيصية؟ وهل مثل هذه المعلومات مطلوبة ومهمة للجمهور في الداخل والخارج؟ وما مدى موثوقية محتويات الملخصات؟

٤- هل بإمكان الحواسب تعزيز عمليتي التدريب والتعديل لدى القائمين على التقدير باستخدام موازين الكفاية اللغوية؟

٥- ما نوع الدعم المطلوب للمشاركين من مدرسين ومتعلمين؟ وهل يمكن أن يكون هذا الدعم بواسطة الحاسوب؟

٦- هل يمكن للحاسوب تعزيز عملية التقييم الذاتي لدى المتعلم؟

٧- هل يمكن للحاسوب تقديم المقومات المطلوبة لاستحداث أساليب اختبارية تتسم بالابتكار مع الاحتفاظ بعنصري الصدق والثبات؟

٨- كيف يمكن الاستفادة من البيانات المجموعة عن طريق الاختبارات المحوسبة في المجالات ذات الصلة؟

٩- ما الذي يمكن أن يتعلمه معدو الاختبارات المحوسبة من التطبيقات "الذكية" للحاسب؟ (Corbel, 1993 : ص ٥٣).

بإلقاء نظرة على الجهود التي أجريت في مجال التقييم اللغوي، تتبين لنا الأفكار التي طرحها كل من "آلدرسون" و"كوربل"، غير أن هذه القضايا المهمة لم تحظ بدراسة موسعة أو دائمة؛ لأن هذا الأمر هو من صميم اختصاص الباحثين في اختبارات اللغة؛ الذين يخطون حاليًا أولى الخطى على طريق فهم الكيفية التي تستطيع طرائق الاختبار الجديدة من خلالها تغيير – وكذلك تطوير – المكونات التي تقاس عن طريق الحاسوب.

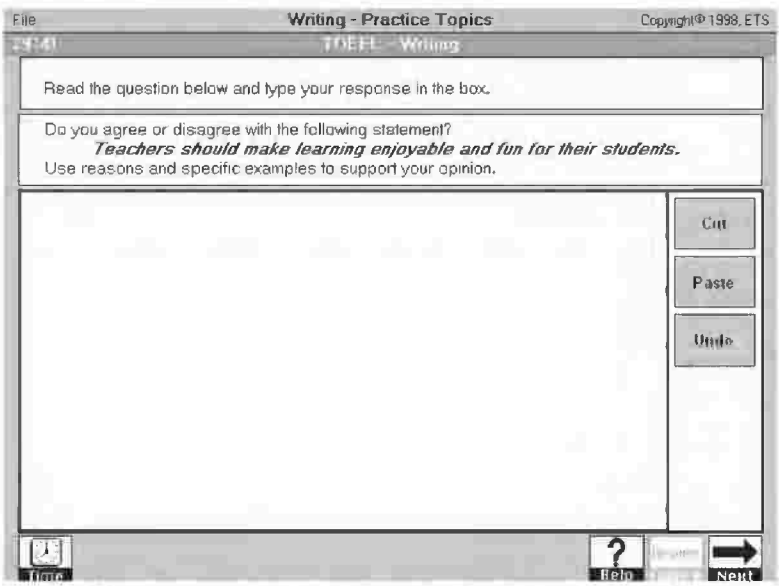
الباحثون في اختبارات اللغة

بعيدًا عن القضايا التي تهم المدرسين ومعدي الاختبارات، فقد تَعرِضُ للباحثين في اختبارات اللغة فرص جديدة للبحث في تلك الاختبارات والتنظير لها مستعينين في ذلك بمستحدثات التقنية. وفي هذا الصدد، قدمت "شابل" (Chapelle, 2003) عرضًا للمجالات التي تهم الباحثين؛ بغية التأكيد على استخدام التقنية وما يتمخض عن ذلك من توسيع نطاق البحث وسبر أغوار الأسئلة البحثية المهمة والإجابة عنها؛ حيث تدعو ثلاثة من تلك المجالات إلى فهم مكونات الاختبارات على نحو أفضل مما هي عليه الحال الآن، في حين يركز مجالان آخران على استكشاف قضايا في الصدق.

توضيح مكونات الاختبارات

تخضع الأسئلة النظرية – التي تتعلق بالمكونات الاختبارية – للتقصي والاستكشاف عندما يواجه مصممو الاختبارات مواقف تستلزم اتخاذ قرارات حول مدى الخيارات المتاحة للمختبرين في أثناء الاختبار، ومن ذلك على سبيل المثال، السماح لهم بمعاودة الاستماع لفقرة صوتية لأكثر من مرة، أو المساعدة في توضيح معاني بعض الكلمات، أو كيفية

التعامل مع الوقت. وإذا ما أتاح الاختبار للمختبرين خيارات كتلك، فإن نتيجة الطالب لن تعكس القدرة على الاستماع فحسب، وإنما ستشمل أيضًا قدرته على استخدام إستراتيجيات مناسبة للتعامل مع تلك الخيارات. وتتطلب هذه المشكلات، التي يواجهها مصممو الاختبارات من الباحث إعادة النظر في إمكانية إضافة عامل الكفاية الإستراتيجية للمختبر ضمن العوامل التي يجب أن يتضمنها الاختبار. فلو نظرنا لاختبار "التوفل" المحوسب كما هو موضح في الشكل رقم (١,٣)، لوجدنا أن خيار المساعدة (Help Option) يظهر في الجانب الأيمن، وبالتحديد في أسفل الشاشة، ويمكن للمختبر اللجوء إليه في أثناء الاختبار لمعرفة كيفية الاستجابة لشكل معين من بنود الاختبار.



الشكل رقم (١,٣). لقطة من أنموذج اختبار "التوفل" المحوسب الخاص بالكتابة.

(Educational Testing Service, 1999)

يتضمن الشكل السابق زرًا يحمل اسم المساعدة "Help" في الجزء السفلي إلى يمين الشاشة، كما أن هناك "المؤقت" الموضح في أعلى يسار الشاشة والذي يمكن إيقافه من

خلال علامة الساعة الموجودة في أسفل يسار الشاشة، وكلها تتيح خيارات يقرر الطالب بنفسه كيفية استخدامها. ولم يتضح بعد إلى أي مدى قد تؤثر هذه الخيارات في أداء المختبرين الذين يستخدمونها، ومن ثم على تحديد قدراتهم اللغوية. فعلى سبيل المثال، قد يلجأ الطلاب للضغط على زر المساعدة في أي وقت في أثناء الكتابة لمعرفة كيفية التعامل مع المفاتيح ذات الصلة، كمفاتيح المسح (Backspace) والحذف (Delete) وبداية السطر (Home) ونهايته (End) ومفاتيح الأسهم وتحريك الصفحة إلى أعلى (Page up) وتحريكها إلى أسفل (Page down) وأخيرًا مفتاح (Enter). ويحسم الوقت الذي قضاه الطالب في المساعدة والتعليمات من وقت الكتابة الفعلي، ومن هنا تتضح لنا مدى أهمية إتقان الطالب لاستخدام تلك الخيارات. وبناء على ما سبق، فإن الدرجة التي يحصل عليها الطالب، تشمل قدرته على الكتابة من جانب، وقدرته على استخدام خيارات المساعدة والتعليمات الحاسوبية من جانب آخر.

وثمة مشكلة جلية أخرى مشابهة تتعلق بقياس فهم المسموع، فالباحثون في اختبارات اللغة يجتهدون أيَّما اجتهاد للإجابة عن أسئلة مثل: هل التعريف النظري لفهم المسموع يأخذ في الاعتبار المعينات البصرية التي تشتمل عليها المادة المسموعة (Gruba, 1997)؟ وكيف يمكن الوصول بالأصالة إلى أفضل صورها في ضوء مجموعة الخيارات المعقدة المحيطة بالأمر؟ فعلى سبيل المثال، يتساءل "كونيام" (Coniam, 2001) عن مدى أصالة اختبار الاستماع المصحوب بالفيديو، ويقدم لذلك مثالاً " بنتيجة اختبار طلاب في هونج كونج حيث " يغلب على الظن بأن المفحوصين في اختبار اللغة الإنجليزية في هونج كونج، أقرب لتعلم الاستماع عبر التلفاز ... منه إلى المذياع" (ص ٣). هذا من جانب، أمَّا من الجانب الآخر، فيصل إلى رأي في نهاية المطاف يرى فيه أن "اختبار الاستماع يجب أن يكون معنيًا في المقام الأول بتقييم مهارة الاستماع"، ولا شك في أن استخدام الفيديو يحول دون تحقيق هذا الهدف. ويردف هذا بقوله: "وفي الوقت الذي

يمكن فيه تعزيز عنصر الأصالة في الاختبار باستخدام الفيديو، فإن عنصر الصدق قد يتأثر سلبًا بهذا الصنيع" (Coniam, 2001: ص ١٢). وعلى هذا فإنه ينبغي أن يكون واضحًا في أذهان الباحثين في اختبارات اللغة، أن هذا الرأي قد يُصاغ في صورة أكثر دقة باعتباره قضية مهمة تخص تعريف المادة المسموعة التي سيقيسها الاختبار، والمجالات التي يحتاج إليها مستخدمو الاختبارات ليكونوا قادرين على استنباط المعنى الذي تمثله الدرجة التي حصل عليها الدارس في الاختبار. أضف إلى ذلك أن وجود الفيديو كأحد العناصر في إعداد الاختبارات، كان هو السبب في طرق تلك القضية — عن المكون الاختباري— بما لها من أهمية معروفة.

تفصيل المكونات الاختبارية

ثمة قضية ثانية تتطرق إلى مكونات الاختبار بأساليب وطرائق جديدة، ألا وهي التعامل مع استجابات الطلاب عبر خوارزميات حاسوبية لرصد الإجابات وتسجيلها. عند وضع برنامج حاسوبي لتحليل لغة المتعلمين وتقويمها، إذ يلزم معدّ الاختبار عندئذٍ اتخاذ قرارات دقيقة بشأن الإجابات الأصوب والأقل صوابًا والتي يمكن أن يقدمها المتعلمون، ولا بد لهذه القرارات من الاعتماد على المكون المطلوب قياسه، غير أن نظرية المكونات الاختبارية المعبر عنها بعبارات عامة مثل "القدرة الكتابية" أو "الثروة اللفظية" لا توفر إرشادًا كافيًا لاتخاذ تلك القرارات بشأن عملية الرصد تلك. ولذلك فإن اختبارات اللغة المحوسبة، تقتضي من الباحث النظر في كيفية استحداث طرائق أو نظريات جديدة للمكونات الاختبارية تكون أكثر تفصيلاً ويتم من خلالها الوصول إلى قرارات دقيقة بشأن استجابات المتعلمين.

تبدو هذه القضية واضحة تماماً مما رأينا من عدم رضا كثير من الباحثين في مجال اختبارات اللغة إزاء نظام كرايْتيرْيَنْ (المعيار) (Criterion) الذي سبق ذكره في مستهل هذا الفصل، إذ إنه استنادًا إلى تقنية معالجة اللغات الطبيعية (NLP) فإن "نظام المعيار"

يُعْمِلُ عنصر التمييز بين المدخلات النصية مسندًا بذلك المسميات إلى طريقة بناء النص في استجابة الطالب، ومحددًا علامات البناء الخطابي (النص) ومفردات المحتوى.كما أن النتيجة التي يقدمها "نظام المعيار" تستمد من تحليل معالجة اللغات الطبيعية: فالتحليل المفرداتيّ، يشير إلى المدى الذي بلغه الطالب في التعامل مع المهمة (الاختبار)، وكذا درجة التعقيد في الاستجابة، أما تحليل الخطاب (discourse analysis) فيقيم مدى التنظيم والتطور، في حين يحكم على جودة التراكيب عن طريق التحليل الخاص بصحتها. وبمعنى آخر، فإن معنى "القدرة على الكتابة" كما يعرفها "نظام المعيار" يستمد من جوانب الكتابات النصية التي يستطيع الحاسوب التعرف عليها: وهي مفردات المحتوى، وعلامات الخطاب، وبعض فئات التراكيب المحددة (انظر مركز خدمة الاختبارات التعليمية Educational Testing Service, 2005a). وعلى الرغم من أن بعض الباحثين الذين يميلون إلى تعريف مهارة الكتابة بصورة أكثر شمولاً وتعقيدًا قد يعمدون إلى انتقاد هذا النظام؛ نظرًا لتعريفه" القدرة على الكتابة تعريفا" تقنيا"، فإننا لا يمكن أن نغفل الإسهامات الملموسة للتقنية في تحديد معنى مهارة الكتابة وتعريفها تعريفا" إجرائيا" دقيقًا.

وحتى كتابة هذه السطور، فإن البحوث القليلة التي تناولت التحليل الحاسوبي للاستجابات (computer-assisted response analysis) قد اصطبغت بصبغة استكشافية تثير أسئلة إضافية أكثر من إجابتها عن أسئلة قائمة (Chapelle, 1993 وHolland, 1994 وConiam, 1998 وCarr, Pan & Xi, 2002). غير أن العمل – وإن كان قليلاً – في هذا المجال المعقد يطرح أسئلة ترمي إلى استكشاف الجوانب التي تكتنف تعريف ماهية المكون: أولا، عندما تُقيم الاستجابات عن طريق التحليل الحاسوبي، يُطلب من الباحثين وضع قواعد لتصحيح الاستجابات بدلاً من بناء قراراتهم فقط على "نعم/لا" فيما يتعلق بمدى مطابقة استجابة الطالب لمفاتيح الإجابة من عدمها. وحتى في

الاستجابات المؤلفة من كلمات فردية في اختبار التتمة المنتظمة (cloze test) فإنها تمثل بيانات يمكن تحليلها بطرق كثيرة مختلفة حسب المطلوب قياسه من خلال الاختبار. وعلى سبيل المثال، برنامج للتصحيح يسجل بيانات الأداء، وهذا البرنامج قادر على تحديد وتسجيل الإنتاج الصحيح وغير الصحيح من حيث: ١- الإملاء. ٢- والتطابق الدلالي والبنائي. ٣- والتراكيب. ٤- وكذا اللواحق الاشتقاقية للكلمات. ٥- واللواحق الصرفية (Chapelle, 2003 – الفصل السادس). و تلك الجوانب الخمس في الاستجابات لأسئلة التتمة المنتظمة المؤلفة من كلمات فردية، قد تتسم بدقة كبيرة أو قليلة نسبةً إلى الجملة وسياق النص محل التقييم. وبناء على ذلك، فإنه – فيما يخص الاستجابة بكلمة واحدة – يمكن لبرنامج تحليلي تقييم ثلاثة مستويات – على الأقل – من الصحة لخمسة مكونات، بحيث يكون مجموع التباين للكلمة الواحدة ١٥ نقطة، علمًا بأنه ينبغي أن يكون هذا التباين متوافقا مع المعنى المقصود من المكون الاختباري. ثانيًا، إذا كانت إجراءات التقييم المذكورة ستؤدي إلى تقييم مجموعة من المكونات (أي إيجاد تقييم مستقل لمستوى الإملاء، وآخر لمستوى دقة المعنى الدلالي) فإنه لا بد من توافر أساليب مقبولة لتقييم مدى ثبات هذه النتائج على نحو متوافق مع تعريف المكون الاختباري. وعلى الرغم من أن هذه الطرائق قد أشبعت بحثا (Embretson, 1985 وAckerman, 1994 على سبيل المثال)، فإنها لم تطبق على أساس نظري لفحص بيانات اختبارات اللغة في ضوء تعريف مفصل للمكون الاختباري.

إعادة النظر في صدق الاختبار

تنبثق قضايا الصدق في الاختبارات أيضًا من وجهة نظر الباحث إزاء الحاجة لفهم أشمل لجوانب القدرة اللغوية التي يمكن قياسها باستخدام الحاسوب. وتنشأ هذه القضايا عندما يتعاطى الباحث مع عنصر الصدق مستخدماً طرائق الارتباط فقط

للبحث في العلاقة بين الاختبارات المقدمة عبر الإنترنت، والاختبارات المقدمة بدون استخدام عنصر التقنية. غير أن هذا الأنموذج يفترض أن كل اختبارات اللغة ينبغي أن تقيس الخصائص اللغوية بطريقة واحدة في مختلف سياقات الاستخدام اللغوي، لكن هذه الفرضية جوبهت بنظرية أخرى ذات صلة بتقييم اللغة لأغراض خاصة (LSP). فإذا كان اختبار الاستماع للغة الإنجليزية في مجال الزراعة – على سبيل المثال – يُتوقع له أن يرتبط ارتباطًا وثيقًا باختبار الاستماع للغة الإنجليزية في مجال علوم الحاسوب، فإن هذا يعني تقويض مفهوم الكفاية اللغوية لأغراض خاصة. واستنادًا إلى الفكرة القائلة بأن اختلاف القدرات أو المهارات بين مختلف استخدامات اللغة يبدو أمرًا طبيعيًا، يوضح "دوقلاس" (Douglas, 1998) أن العوامل الأخرى (أي: تأثير الطريقة المستخدمة) لا ينبغي اعتبارها بالضرورة مصادر للأخطاء فيقول:

> بدلاً من السعي إلى الحد من تأثير الطريقة المستخدمة ... نحتاج إلى استغلال ذلك التأثير من خلال تصميم اختبارات تستهدف فئات بعينها – أي اختبارات تشتمل على: تعليمات، ومحتوى، وتخصص، ولغة توجه كلها إلى هذه الفئات (Douglas, 1998: ص١٥٣).

وإذا كان اختبار اللغة الإنجليزية لطلاب الكليات والجامعات، سيتطلب من الطلاب استخدام "الفأرة" لاختيار الاستجابات في اختبار لمهارة الاستماع – على سبيل المثال – فإن السؤال الذي ينبغي طرحه – من وجهة نظر "دوقلاس" هو تقييم هذه الطريقة لمهارات ذات صلة، ولا ينبغي أن يكون السؤال عما إذا كانت تلك الطريقة سببًا في ضعف أو قوة الارتباط باختبار آخر. إن اختبارات اللغة لأغراض خاصة، تحتوي على حالات واضحة لما ينبغي أن تشتمل عليه المكونات اللغوية، أما التقنية فتقدم حالة أكثر غموضا، بل وتضليلا" في بعض الأحيان. ومن المعلوم أن الحاسوب هو أداة تقدم من خلالها الاختبارات الموضوعة لقياس الخصائص اللغوية العامة (على

سبيل المثال: القواعد وفهم المقروء) ولكنه يحتوي على عوامل قد تؤثر في اللغة التي نرغب في تقييمها. ولكن التقنية، في الوقت ذاته، تؤدي إلى إحداث تغييرات فعلية في طريقة استخدام اللغة، ومن ثم في المهارات اللازمة لاستخدامها (Crystal, 2001 على سبيل المثال). وبذلك نجد أن التعقيد الذي يكتنف قضايا الصدق في التقييم اللغوي المحوسب، إنما يقف عائقًا أمام استحداث اتجاه مبسط أو محدد المعالم تجاه هذه القضية. ومن ثمّ، يواجه الباحثون تحديًا في استقصاء الافتراضات المعمول بها سابقًا والتحقق من صحتها، إلى جانب استحداث بدائل أخرى أكثر ثباتًا.

ربط الاختبارات المحوسبة ببحوث اكتساب اللغة الثانية

في ضوء ما أشار إليه كل من "آلدرسون" و"كوربل" في استكشافهما لإمكانات الحاسوب وآفاقه، فإن التقنية تسمح لقياس مكونات اختبارية مختلفة؛ لاسيما تلك المكونات المقيّمة عبر عنصر الحفاظ على الوقت مثلاً. وفي بحوث اكتساب اللغة الثانية (SLA)، جاء تفسير وقت التأخير قبل أن يستجيب المتعلم لأحد الأسئلة على أنه مؤشر للأتمتة (Automaticity) في الأنشطة البحثية (Hagen, 1994) أو على أنه تحضير مسبق (Advance Preparation) في الأنشطة المصحوبة بتعليمات (Jamieson & Chapelle 1987). وفي مثل هذه الدراسات يقدم التقييم اللغوي المحوسب وسيلة لقياس المكونات الاختبارية محل الاهتمام لدراسة اكتساب اللغة واستخدامها، الأمر الذي لا يُتاح من خلال أداء الدارسين في الاختبارات الورقية. بالإضافة إلى ما سبق، يمكن تقييم هذه العمليات في أثناء مشاركة الدارسين في أنشطة تعليمية عادية عبر الإنترنت، مما يعني فتح آفاق أرحب وأرجى لتقييم بعض الإستراتيجيات والأساليب المستخدمة تقييمًا تشخيصيًا عبر الإنترنت. وعلى الرغم من اختلاف مثل هذه التقييمات عن أوجه الحكم على صحة (قواعد اللغة) المعروفة، وعن التقارير الذاتية وغيرها من الأساليب المستخدمة في بحوث اكتساب اللغة الثانية، فإنها تثير أسئلة مهمة بشأن صحة

الاستنتاجات في الأنشطة البحثية المتعلقة باكتساب اللغة الثانية. وفي هذا السياق، نرى أن "هيقل هايمر" و"شابل" (Hegelheimer and Chapelle, 2000) – على سبيل المثال – يريان بأن الضغط على الفأرة طلبًا لتعريفات للمفردات، قد يؤخذ على أنه مؤشر على القدرة على الملاحظة، والتي تعد مكوناً مهماً في نظريات الاكتساب في بحوث اكتساب اللغة الثانية، لكن في الوقت ذاته يثير هذا الاستنتاج أسئلة كثيرة بشأن الطرائق المقبولة لتبرير مثل هذه الاستنتاجات.

استقصاء ما يترتب على الاختبارات المحوسبة

منذ أن بدأ استخدام التقنية في التقييم اللغوي، والباحثون يودون معرفة ما إذا كان تخوف الدارسين أثناء الاختبار من استخدام الحاسوب قد يؤثر في أدائهم اللغوي أم لا. وقد طُرح منذ ذلك الحين سؤال حول إمكانية توثيق الآثار السلبية للحاسوب، وكذا كيفية الحد منها قدر الإمكان إن وجدت. غير أن هذا السؤال يفقد أهميته بمضي الوقت؛ لأن قدرة الطلاب على التعامل بكفاءة مع تقنية الحاسوب تتزايد يومًا بعد آخر، بل إن لدينا اليوم طلاباً يعبرون عن قلقهم إزاء فكرة الإجابة عن الاختبار النهائي بالكتابة عبر أداة أخرى غير الحاسوب. من ناحية أخرى، نجد أننا في الوقت ذاته، لا نعرف إلا أقل القليل عن مدى تأثير قلق الطلاب إزاء استخدام الحاسوب على أدائهم. لذا فإن مصممي الاختبارات في الوقت الحاضر، يعملون على مساعدة الطلاب الذين يراودهم الشعور بالقلق لدى استخدام التقنية عند أداء الاختبار. ولعل أفضل الأمثلة في هذا الصدد يتمثل في التدريب الخاص الذي يسبق اختبار "التوفل" المحوسب (Jamieson, Kirsch, Taylor & Eignor, 1999). حيث يتيح هذا التدريب الفرصة لجميع المتعلمين للاطلاع على الوظائف الكتابية والتحريرية ووظائف التحكم في الحاسوب قبل البدء في الاختبار؛ وذلك بغية الحد – قدر المستطاع – من احتمال التباين في إجابات الطلاب الذي يكون مرده جزئياً إلى ألفتهم أو عدم ألفتهم باستخدام الحاسوب.

تؤكد التقنية الحاجة إلى أن يقوم الباحثون بدراسات على نحو يكشف عن الآثار المترتبة على الاختبارات المحوسبة، تمامًا كما فعل بعض الباحثين مع الاختبارات التقليدية التي تستخدم الورقة والقلم (انظر مثلاً Alderson&Hamp-Lyons, 1996). غير أن هناك منظوراً آخر لا يقل أهمية عما سبق، يتمثل في تصميم تدريبات مسبقة على الاختبارات، من شأنها تحفيز المتعلمين ومدرسيهم وتشجيعهم لاستخدام التقنية في التعلم والتعليم. وبمعنى آخر، فإن الاختبار المحوسب قد ينطوي على نوع من التأثير الإيجابي في اكتساب اللغة إذا أخذنا في الحسبان الفوائد التي يرى الكثيرون ارتباطها بالتعامل المنتظم مع التقنية في أثناء تعلم اللغة الثانية. ويرى كذلك كثير من المدرسين والباحثين، أنه ينبغي للطلاب الوصول إلى المواد التعليمية عبر الإنترنت؛ لأن هذا الأمر من شأنه زيادة التعامل مع الحاسوب واكتساب المزيد من الجوانب المطلوبة لإتقان اللغة المتعلَّمة، فهذا "وارشور" (Warschauer, 1998) – على سبيل المثال – يرى أن إتقان الأمرين (التعامل مع الحاسوب واكتساب المزيد لإجادة اللغة المتعلمة) يشكل عنصرًا أساسيًا للنجاح، وهو ما يثير أسئلة بحثية مهمة من بينها "كيف تصاغ تدريبات اللغة واكتساب مبادئ الحاسوب من خلال عوامل اجتماعية ومؤسسية أوسع نطاقًا؟ وما الذي تعنيه هذه الممارسات الجديدة من منظور المتعلم؟" (Warschauer, 1998: ص ٧٦٠). ومن الممكن أن تتناول دراسة الآثار المترتبة على استخدام التدريبات الحاسوبية في الاختبارات مثل هذه الأسئلة، علمًا بأن نظرية صدق الاختبار لن تحث الباحثين في المستقبل على توثيق التبعات السلبية للاختبارات المحوسبة فحسب، بل كذلك الآثار الإيجابية المحتملة لتلك الاختبارات.

التطلعات إلى نجاح اختبارات اللغة المحوسبة

في ضوء الفرص والتحديات الصعبة التي تبدو التقنيات قادرة على توفيرها لمجال التقييم اللغوي، فإنه من المطمئن معرفة أن بعض الباحثين درسوا التقنيات وتأثيرها في

التقييم اللغوي منذ ما يزيد على عشرين عامًا، حيث عُرض الموضوع على بساط البحث كمحور رئيس لمؤتمر أبحاث اختبارات اللغة الذي عقد عام ١٩٨٥م ونشرت بعض أوراقه العلمية في سجل للمؤتمر حرره "ستانسفيلد" عام ١٩٨٦م (Stansfield, 1986). وظهرت هذه المجموعة من المقالات في وقت لم يكن فيه إلا قلة قليلة من المهتمين باختبارات اللغة كانوا هم الذين يستطيعون التعامل مع التقنية. وقد استشرف ذلك الإصدار الكثير من القضايا التي ستشغل أذهان الباحثين والمشتغلين باختبارات اللغة على مدار العقد المقبل على الأقل. وقد أشارت بعض الأبحاث الواردة في سجل المؤتمر، إلى مجالات الاهتمام التي تشغل أذهان مدرسي اللغة، ومعدي الاختبارات، والباحثين كما أسلفنا، لكن هذه المجالات لم تخضع لبحث شامل ودقيق، ولم يصل فيها الباحثون إلى حلول نهائية. لقد رأينا في هذا الفصل أن الاشتغال بمسار التقنية الذي بدأ في مؤتمر عام ١٩٨٥م — وما تلاه من تناول في مختلف كتب سلسلة كامبردج للتقييم اللغوي — كل ذلك جدير بالدراسة بكونه موضوعًا مستقلاً، فهو موضوع ينطوي على آفاق جديدة للمشتغلين باختبارات اللغة ، ومن ضمنهم المدرسون والباحثون والمتعلمون.

يبحث الفصل التالي في مظاهر الاختلاف بين طرائق الاختبارات التي تستخدم فيها التقنية، والطرائق التقليدية المعمول بها في الماضي والتي كانت تعتمد على الورقة والقلم، والصوت، والمقابلات الشخصية الشفهية. وتشكل هذه الاختلافات أساسًا ننطلق منه للنظر في القضايا التي تحظى باهتمام المدرسين ومعدي الاختبارات والباحثين، والتأكد ما إذا كانت هذه الطرائق قادرة على الإسهام إيجابًا في ترسيخ عنصر الصدق في الاستنتاجات المستمدة من التقييمات واستخداماتها. أما الفصل الثالث فسيتناول احتمال إسهام اختبارات اللغة المحوسبة في تشويه أو تغيير القدرة اللغوية التي تُقاس بالاختبار، بالقدر الذي لا يمكن معه اعتبار معدل أداء الدارس في الاختبار دليلاً على قدرته اللغوية.

وسيناقش هذا المهدد الناتج عن الاستعانة بالتقنية من خلال النظر في الأدلة النظرية والتجريبية التي تخص مفهوم أداء الطلاب في اختبار اللغة المحوسب.

يتناول الفصل الرابع وصفاً للأدوات البرمجية المستخدمة في إعداد اختبارات اللغة المحوسبة، ويبين بعض المواضع المهمة التي نحتاج إليها عند إعداد الاختبارات المعتمدة على التقنية، والتي تختلف عن غيرها من صور الاختبارات الأخرى، أي أنَّ الفصل يتناول الخيارات المتاحة أمام معدي الاختبارات، ويصف تلك البرمجيات في ضوء استخدامها في اختبارات اللغة. أما الفصل الخامس فيبحث في الكيفية المثلى لتقييم اختبارات اللغة المحوسبة، وإذا اتفقنا أن هذا النوع من الاختبارات يختلف عن الاختبارات التقليدية التي تعتمد على الورقة والقلم، فهل ينبغي تقييمه استنادًا إلى معايير تختلف عن المعايير المستخدمة في الاختبارات الأخرى؟ ثم نمضي بين دفتي الكتاب إلى الفصل السادس، وفيه نتناول مقدار إسهام التغييرات في الاختبارات المعتمدة على التقنية في تشكيل الثورة التي تنبأ بها بعضهم في عالم الاختبارات. ونحن نرى أن تلك التغييرات تنم عن تطور أكثر مما تنم عن ثورة حقيقية في هذا الصدد، وأن تلكم التغييرات التي تنم عن ثورة حقيقية، تحتاج إلى أن تكون مدفوعة من خلال التطورات التي تطرأ على فهمنا لاستخدام اللغة واكتساب اللغة الثانية.

ويهدف الكتاب عموماً إلى إلقاء الضوء على الآفاق الجديدة للتقنية في مجال التقييم اللغوي، وما تنطوي عليه من قضايا لذا يركز في المقام الأول على مدى التغيير الناتج عن تقديم الاختبارات عبر الإنترنت واستخدام التقنيات التفاعلية، ومعالجة اللغات الطبيعية، وأثر ذلك في تجربة الطلاب والمستخدمين من المعلمين، وكذا قضايا التقييم التي تهم معدي اختبارات اللغة والباحثين في هذا المجال. تلكم هي الجوانب التي حظيت باهتمام أقل إذا ما قورنت بالاختبارات الحاسوبية المتكيفة، والتحليل الإحصائي لبيانات الاختبارات العددية، وهما مجالان يمثلان تطورات مهمة طرأت على التقييم اللغوي بفضل استخدام التقنية.