

أنواع الاختبارات القائمة على أساس الاختيار

تأليف

مدرس الدكتور محمد لوتي

أكتوبر 1, 2025

اقتبس من هذا المقال

مدرس الدكتور محمد لوتي (2025). أنواع الاختبارات القائمة على أساس الاختيار. عرب سايكولوجي. تم الاسترجاع من <https://arabpsychology.com/?p=28089>

أنواع الاختبارات القائمة على أساس الاختيار في علم النفس

مقدمة: رحلة إلى عالم تقييم الاختيار المتعدد

يُعد التقييم حجر الزاوية في أي عملية تعليمية أو مهنية فعالة، فهو ليس مجرد إجراء روتيني، بل هو بوصلة توجه مسار التعلم والتطوير، ومرآة تعكس مدى تحقيق الأهداف المنشودة وتقييم الكفاءات والمهارات المكتسبة. في هذا السياق الديناميكي، تتنوع أدوات التقييم وأساليبه لتلبي الأغراض المختلفة وتناسب المستويات المعرفية المتعددة التي يُراد قياسها، من أبسط الحقائق إلى أعقد عمليات التفكير النقدي. تهدف هذه التنوعية إلى توفير صورة شاملة ودقيقة لأداء الأفراد، مما يُمكن من اتخاذ قرارات مستنيرة تتعلق بالتحصيل الأكاديمي، التطور المهني، أو حتى التوظيف.

من بين الطيف الواسع لأساليب التقييم، تبرز الاختبارات القائمة على الاختيار (Selection-Based Tests) كواحدة من أكثر الأدوات شيوعاً وانتشاراً، نظراً لمرونتها وكفاءتها. تتميز هذه الاختبارات بكونها تتطلب من المفحوصين اختيار الإجابة الصحيحة أو الأنسب من بين مجموعة من البدائل المتاحة، بدلاً من إنشائها بأنفسهم. لقد أحدث هذا النهج، وبخاصة أسئلة الاختيار من متعدد (Multiple-Choice Questions - MCQs)، ثورة حقيقية في مجال القياس والتقييم، وذلك لما تتمتع به من موضوعية عالية في التصحيح وكفاءة ملحوظة في تغطية محتوى واسع في وقت قصير نسبياً، كما أكد العديد من الخبراء في هذا المجال مثل بوبهام (Popham, 2017). ومع ذلك، فإن تحقيق الفعالية القصوى لهذه الاختبارات لا يعتمد فقط على مجرد تطبيقها، بل يتوقف بشكل حاسم على جودة تصميمها وبنائها الدقيق.

إن اكتساب فهم عميق ومفصل لأنواع هذه الاختبارات، وخصائصها المميزة، ومزاياها الجوهرية، وعيوبها المحتملة، بالإضافة إلى المبادئ التوجيهية السليمة لبنائها، يُعد أمراً بالغ الأهمية لكل من يُعنى بالتقييم. يشمل ذلك المعلمين الذين يسعون لقياس فهم طلابهم، والمطورين التربويين الذين يصممون المناهج، وأخصائيي القياس الذين يضمنون دقة وموثوقية الأدوات، وكذلك أصحاب العمل الذين يستخدمون هذه الاختبارات لتقييم الكفاءات. عند استخدامها بشكل صحيح ومتقن، توفر هذه الاختبارات بيانات قيمة يمكن الاعتماد عليها لاتخاذ قرارات مؤثرة، سواء كانت تتعلق بتحديد نقاط ضعف الطلاب لتعديل استراتيجيات التدريس، أو في عمليات الفحص الأولية للمتقدمين لوظائف مرموقة، حيث تُصبح الكفاءة والموضوعية عاملين حاسمين (Downing & Haladyna, 2006). يهدف هذا المقال إلى تقديم تحليل شامل ومتعمق لهذه الاختبارات، مع التركيز على أسئلة الاختيار من متعدد كنموذج رائد، وذلك لتمكين القارئ من تسخير قوتها الكامنة بفعالية ومسؤولية.

1. التعريف الجوهرى للاختبارات القائمة على الاختيار

تُعرّف الاختبارات القائمة على الاختيار بأنها فئة محددة من أدوات التقييم التي تضع أمام المفحوص مجموعة من الإجابات المحتملة أو البدائل لسؤال، عبارة، أو مشكلة معينة، وتتطلب منه تحديد الخيار الذي يمثل الإجابة الصحيحة أو الأفضل. السمة الجوهرية التي تميز هذه الاختبارات عن غيرها هي أن المفحوص لا يقوم بإنشاء الإجابة من ذاكرته أو بناء استجابته الخاصة، كما هو الحال في الأسئلة المقالية أو الإجابات القصيرة، بل يعتمد على قدرته على التعرف على الإجابة الصحيحة أو تمييزها من بين الخيارات المعروضة أمامه. هذا التركيز على التعرف بدلاً من الاستدعاء يُشكل أساساً منهجياً وفلسفياً يميز هذه الاختبارات ويحدد طبيعة المهارات المعرفية التي تقيسها، مما يجعلها أداة فعالة لتقييم نطاق واسع من المعارف والقدرات.

يكمن المبدأ الأساسي وراء هذه الاختبارات في تقديم سيناريو محدد أو مشكلة، ثم تتبعها بخيارات متعددة، واحد منها فقط هو الأنسب أو الأدق علمياً. هذا الأسلوب يقلل من تأثير عوامل مثل القدرة على التعبير الكتابي أو سرعة الكتابة، ويركز بدلاً

من ذلك على القدرة على معالجة المعلومات، التمييز بين الصواب والخطأ، وتطبيق المفاهيم في سياقات معينة. تُستخدم هذه الاختبارات على نطاق واسع في مجالات مثل علم النفس التربوي، القياس النفسي، والتقييمات المهنية، لكونها توفر طريقة موحدة وموضوعية لجمع البيانات حول الأداء الفردي.

2. السياق التاريخي والتطور

تعود جذور الاختبارات القائمة على الاختيار، وبخاصة أسئلة الاختيار من متعدد (MCQs)، إلى أوائل القرن العشرين مع ظهور حركة القياس التربوي الحديثة والحاجة إلى أدوات تقويم أكثر كفاءة وموضوعية. يُنسب الفضل في تطوير أول اختبار اختيار من متعدد رسمي إلى فريدريك جاي كيلبي (Frederick J. Kelly) في عام 1914، كجزء من اختبار القدرة العقلية باللغة الإنجليزية في جامعة كانساس. جاء هذا التطور استجابة للحاجة المتزايدة لتقييم أعداد كبيرة من الطلاب بكفاءة، خاصة في ظل التوسع في التعليم العام خلال تلك الفترة.

شهدت أسئلة الاختيار من متعدد انتشاراً واسعاً خلال الحرب العالمية الأولى، عندما استخدمت الجيوش الأمريكية اختبارات "ألفا وبيتا" لتصنيف المجندين وتوزيعهم على المهام المختلفة بناءً على قدراتهم المعرفية. كانت هذه الاختبارات تعتمد بشكل كبير على بنود اختيار من متعدد لضمان سرعة التصحيح والموضوعية. في العقود اللاحقة، استمر التطور النظري والتطبيقي لهذه الاختبارات، مع مساهمات بارزة من باحثين مثل إيبيل (Ebel) وغرونلوند (Gronlund) وهالادينا (Haladyna)، الذين وضعوا أسساً علمية لتصميم البنود وتحليلها، مما أدى إلى تحسين جودتها وقدرتها على قياس مستويات معرفية أعلى.

أصبحت الاختبارات القائمة على الاختيار جزءاً لا يتجزأ من أنظمة التقييم العالمية، من امتحانات القبول الجامعي مثل SAT و GRE، إلى الاختبارات المهنية، والتقييمات المدرسية الشاملة. هذا التطور التاريخي يعكس قدرة هذه الأدوات على التكيف مع الاحتياجات المتغيرة للتقويم، مع السعي المستمر لتحقيق التوازن بين الكفاءة، الموثوقية، والصدق في القياس.

3. الأشكال الرئيسية للاختبارات القائمة على الاختيار

يشتمل تصنيف الاختبارات القائمة على الاختيار على عدة أشكال شائعة ومتنوعة من بنود الاختبار التي تُستخدم على نطاق واسع في البيئات التعليمية والمهنية. كل شكل من هذه الأشكال له خصائصه الفريدة، ومجالات استخدامه المثلى، وتحدياته الخاصة في التصميم. فهم هذه الأشكال المتنوعة يُمكن واضعي الاختبارات من اختيار الأداة الأنسب لهدف التقييم المحدد والمحتوى المراد قياسه، مما يضمن الحصول على بيانات تقويمية أكثر دقة وموثوقية.

أسئلة الاختيار من متعدد (Multiple-Choice Questions - MCQs)

تُعد أسئلة الاختيار من متعدد هي الشكل الأكثر شيوعاً ودراسة ضمن الاختبارات القائمة على الاختيار، وتُعتبر بحق الركن الأساسي في العديد من أنظمة التقييم الحديثة. يتكون كل سؤال من هذا النوع عادةً من ثلاثة مكونات أساسية تعمل معاً لقياس فهم المفحوصين. أولاً، هناك "الجذع" (Stem)، وهو الجزء الذي يطرح المشكلة، السؤال، أو العبارة غير المكتملة التي يجب على المفحوص الاستجابة لها. ثانياً، توجد مجموعة من "البدايل" (Alternatives)، وهي الخيارات المتعددة التي تُقدم للمفحوص. ضمن هذه البدايل، يُوجد بديل واحد يمثل الإجابة الصحيحة أو "المفتاح" (Key)، وهو الحل الأمثل أو الأصح للمشكلة المطروحة. أما البدايل الأخرى، فتُعرف بـ "المموهات" أو "المشتتات" (Distractors)، وهي إجابات خاطئة أو أقل صحة، تُصمم بعناية لتكون جذابة للمفحوصين الذين يمتلكون فهماً ناقصاً أو مفاهيم خاطئة حول المادة (Haladyna, Downing, & Rodriguez, 2002). تكمن سر قوة هذا النوع في مرونته وقدرته على قياس مجموعة واسعة من

المخرجات التعليمية، تتجاوز مجرد تذكر الحقائق البسيطة، لتصل إلى الفهم، التطبيق، والتحليل.

أسئلة الصواب والخطأ (True/False Items)

تمثل أسئلة الصواب والخطأ شكلاً آخر من الاختبارات القائمة على الاختيار، وتُعرف ببساطتها الظاهرية وسهولة بنائها وتصحيحها. في هذا النوع، تُقدّم للمفحوص عبارة أو جملة خبرية، ويُطلب منه تحديد ما إذا كانت هذه العبارة صحيحة أم خاطئة. على الرغم من أن هذا التنسيق يبدو سهل الإعداد، إلا أنه غالباً ما يواجه انتقادات كبيرة بسبب عدة عوامل تحد من فعاليته كأداة تقييم شاملة. من أبرز هذه الانتقادات هي الاحتمالية العالية للتخمين، حيث تصل نسبة التخمين الصحيح إلى 50%، مما يقلل من دقة القياس وقد يؤدي إلى نتائج لا تعكس الفهم الحقيقي للمادة (Ebel & Frisbie, 1991). كما أنها محدودة في قياس مستويات التفكير العليا، حيث تركز غالباً على تقييم المعرفة بالحقائق البسيطة أو المفاهيم الأساسية.

أسئلة المزوجة أو المطابقة (Matching Items)

تُقدّم أسئلة المزوجة أو المطابقة شكلاً مميزاً وفعالاً ضمن الاختبارات القائمة على الاختيار، وهي مصممة خصيصاً لقياس القدرة على تحديد العلاقات والارتباطات بين مجموعتين من المعلومات المتجانسة. يتكون هذا النوع من قائمتين رئيسيتين: القائمة الأولى (Premise List) التي تحتوي على مجموعة من العناصر، مثل مصطلحات، أسماء، تواريخ، أو أسباب؛ والقائمة الثانية (Response List) التي تحتوي على مجموعة من الإجابات المحتملة، مثل تعريفات، أحداث، نتائج، أو خصائص. يُطلب من المفحوص ربط كل عنصر في القائمة الأولى بالعنصر المناسب له من القائمة الثانية. تتميز بفعاليتها العالية في تغطية كمية كبيرة من المحتوى المترابط في مساحة محدودة نسبياً من الاختبار، وتقلل من تأثير التخمين مقارنة بأسئلة الصواب والخطأ.

4. تشريح سؤال الاختيار من متعدد ومثال عملي

لفهم كيفية عمل أسئلة الاختيار من متعدد وتقدير إمكاناتها الكبيرة وقيودها المحدودة، من الضروري التعمق في تحليل مكوناتها الأساسية التي تتفاعل معاً لتشكل أداة قياس فعالة. يتألف سؤال الاختيار من متعدد النموذجي من جزأين رئيسيين هما: الجذع (Stem) والبدائل (Alternatives). هذه البدائل تنقسم بدورها إلى الإجابة الصحيحة والمموهات. كل مكون من هذه المكونات له دور حيوي في تحديد جودة السؤال وقدرته على قياس المعرفة أو المهارة المستهدفة بدقة. إن الإتقان في تصميم كل جزء من هذه الأجزاء يُعد مفتاحاً لإنشاء اختبارات اختيار من متعدد ذات جودة عالية يمكن الاعتماد عليها في اتخاذ القرارات التقييمية الهامة.

الجذع (Stem): يمثل الجذع جوهر سؤال الاختيار من متعدد، فهو الجزء الذي يطرح المشكلة أو السؤال أو يقدم العبارة غير المكتملة التي يجب على المفحوص الاستجابة لها. يجب أن يتحلّى بالوضوح والدقة، وأن يكون مفهوماً بذاته، مما يعني أن المفحوص يجب أن يستطيع فهم المشكلة المطروحة تماماً دون الحاجة إلى قراءة البدائل أولاً.

البدائل (Alternatives): تمثل مجموعة الخيارات التي تُقدّم للمفحوص ليختار من بينها الإجابة الصحيحة. تتضمن البدائل عادةً إجابة صحيحة واحدة، تُعرف بـ المفتاح (Key)، ومجموعة من الإجابات الخاطئة المصممة بعناية، والتي تُسمى المموهات أو المشتتات (Distractors). إن جودة هذه البدائل، وخاصة المموهات، هي العامل الأكثر أهمية في تحديد فعالية سؤال الاختيار من متعدد وقدرته على قياس ما يُفترض به قياسه بدقة. يجب أن تكون المموهات "معقولة" (plausible)، أي تبدو كإجابات محتملة ومقبولة للمفحوصين الذين لم يتقنوا المادة بشكل كامل.

مثال عملي: تطبيق مبدأ التعزيز الإيجابي

لنفترض أننا نريد تقييم فهم طالب في علم النفس التربوي لمفهوم التعزيز الإيجابي. يمكننا صياغة سؤال الاختبار من متعدد كالتالي:

الجدع: يُلاحظ معلم أن أحد طلابه يتردد في المشاركة في الأنشطة الصفية. بعد كل مرة يشارك فيها الطالب، يقوم المعلم بالثناء عليه وتقديم ملصق مكافأة. أي من المفاهيم التالية يُمثل الإجراء الذي اتخذته المعلم لزيادة احتمالية مشاركة الطالب في المستقبل؟

البدائل:

العقاب الإيجابي

التعزيز الإيجابي (الإجابة الصحيحة - المفتاح)

التعزيز السلبي

العقاب السلبي

في هذا المثال، الجذع يقدم سيناريو واضحاً يتطلب تطبيق مفهوم نفسي. البدائل الثلاثة الأخرى (العقاب الإيجابي، التعزيز السلبي، العقاب السلبي) تعمل كموهات معقولة. الطالب الذي يفهم مبادئ الإشراف الإجرائي سيختار "التعزيز الإيجابي" لأنه يتضمن إضافة محفز مرغوب فيه (الثناء والملصق) لزيادة سلوك مرغوب فيه (المشاركة). هذا المثال يوضح كيف يمكن لسؤال الاختبار من متعدد أن يقيس القدرة على التطبيق وليس مجرد التذكر.

5. المستويات المعرفية المستهدفة: من التذكر إلى التفكير النقدي

بالإضافة إلى التصنيفات البنائية، يمكن أيضاً تصنيف أسئلة الاختبار من متعدد بناءً على المستوى المعرفي الذي تهدف إلى قياسه، وذلك بالاستناد إلى تصنيفات الأهداف التربوية المعترف بها مثل تصنيف بلوم المعدل (Anderson & Krathwohl, 2001). هذا التصنيف يُمكن واضعي الاختبارات من بناء تقييمات متوازنة تغطي طيفاً كاملاً من القدرات العقلية، من الاسترجاع البسيط للمعلومات إلى العمليات المعرفية المعقدة التي تتطلب التفكير النقدي والإبداعي.

التذكر (Remembering): تقيس هذه الأسئلة القدرة الأساسية على استدعاء أو التعرف على معلومات محددة تم تعلمها مسبقاً، مثل الحقائق والمصطلحات والتعريفات. مثال: "ما هو اسم العالم الذي اقترح نظرية التعلم بالملاحظة؟".

الفهم (Understanding): تتطلب هذه الأسئلة من المفحوصين إظهار فهمهم للمعلومات من خلال تفسيرها، تلخيصها، شرحها بكلماتهم الخاصة، أو استنتاج المعنى الضمني. مثال: "أي من العبارات التالية يمثل أفضل تلخيص لمفهوم التنافر المعرفي؟".

التطبيق (Applying): تقيس أسئلة التطبيق القدرة الحيوية على استخدام المفاهيم، المبادئ، القواعد، أو الإجراءات في مواقف جديدة أو مختلفة. مثال: "إذا كان طفل يبكي للحصول على الحلوى، والأم تتجاهل بكاءه حتى يتوقف، فما هو مبدأ التعلم الذي تطبقه الأم؟".

التحليل (Analyzing): تتطلب هذه الأسئلة من المفحوصين تقسيم المعلومات إلى مكوناتها الأساسية، وتحديد العلاقات بين

الأجزاء، والتعرف على الافتراضات الضمنية، والتمييز بين الحقائق والآراء. مثال: "ما هو الافتراض الأساسي الذي تقوم عليه حجة النظرية السلوكية حول اكتساب اللغة؟".

التقويم (Evaluating) والابتكار (Creating): يُعتبر قياس هذين المستويين العلويين من تصنيف بلوم باستخدام أسئلة الاختبار من متعدد تحدياً كبيراً (Scully, 2017). بينما يمكن تصميم أسئلة معقدة تتطلب حكماً نقدياً (تقويمياً)، فإن قياس القدرة على الإنشاء أو الابتكار الأصيل غالباً ما يتطلب أساليب تقويم بديلة. ومع ذلك، يمكن لأسئلة الاختبار من متعدد المصممة بعناية فائقة أن تقترب من تقييم جوانب من هذه المهارات العليا، مثل التعرف على أفضل نقد لحجة معينة أو اختيار أفضل تصميم من بين عدة خيارات.

6. الأهمية والتأثير في علم النفس والتعليم

تتمتع الاختبارات القائمة على الاختيار، وبشكل خاص أسئلة الاختبار من متعدد، بمجموعة فريدة من المزايا التي رسخت مكانتها كأدوات تقويم لا غنى عنها في مختلف المجالات التعليمية والمهنية، مما يعكس تأثيرها العميق. تُقدم هذه الاختبارات حلولاً عملية وفعالة لتحديات القياس النفسي والقياس التربوي، ولكنها أيضاً تحمل قيوداً وتحديات يجب إدراكها.

المزايا الرئيسية:

الموضوعية الفائقة في التصحيح: تُعد هذه الميزة الأبرز، حيث يمكن تصحيح الاختبارات بدقة متناهية وسرعة فائقة، سواء يدوياً أو آلياً. هذا يزيل أي تحيز محتمل ويضمن الاتساق المطلق في الدرجات، وهو أمر بالغ الأهمية في الاختبارات عالية المخاطر.

الكفاءة وتغطية المحتوى الشاملة: تُتيح أسئلة الاختبار من متعدد الإجابة على عدد كبير من البنود في وقت محدود، مما يُمكن من تغطية نطاق أوسع من المحتوى ويعزز من الصدق المحتوي للاختبار، مما يجعله ممثلاً بشكل أفضل للمجال المعرفي المستهدف.

الموثوقية العالية: بفضل موضوعية التصحيح وإمكانية تضمين عدد كبير من البنود، تميل هذه الاختبارات إلى تحقيق درجات موثوقة عالية، أي اتساق الدرجات عبر تطبيقات مختلفة، مما يعزز الثقة في النتائج.

إمكانية التحليل الإحصائي للبنود: تُقدم هذه الاختبارات ميزة فريدة في إمكانية التحليل الإحصائي المفصل لكل بند، مما يساعد في تقييم جودته وفعاليته، وتحديد البنود الضعيفة، وتشخيص مجالات سوء الفهم لدى الطلاب، وبالتالي تحسين جودة التعليم والتقييم.

القيود والتحديات:

عامل التخمين: يُعد عامل التخمين أحد أبرز الانتقادات، حيث يمكن للمفحوصين الحصول على درجات معينة عن طريق التخمين العشوائي، مما يقلل من دقة القياس وقد يشوه الدرجات، خاصة في الاختبارات القصيرة.

صعوبة قياس مستويات التفكير العليا والمهارات الإنتاجية: تُنتقد هذه الاختبارات لتركيزها المفرط على التذكر والتعرف بدلاً من مهارات التفكير العليا مثل التحليل النقدي، التقويم، حل المشكلات المعقدة، أو الإبداع. كما أنها لا تقيس القدرة على تنظيم الأفكار والتعبير عنها.

احتمالية وجود تلميحات غير مقصودة (Cueing): قد تحتوي صياغة الجذع أو البدائل على تلميحات غير مقصودة يمكن للمفحوص استغلالها للوصول إلى الإجابة الصحيحة دون فهم حقيقي للمادة، مما يقلل بشكل كبير من صدق الاختبار.

استهلاك الوقت والجهد في البناء الجيد: على الرغم من سهولة تصحيحها، فإن بناء أسئلة اختيار من متعدد جيدة، خاصة تلك التي تقيس مستويات التفكير العليا وتتضمن مموهات معقولة، هو عملية تتطلب وقتاً وجهداً كبيرين وخبرة متخصصة.

التأثير السلبي المحتمل على التدريس والتعلم (Negative Backwash Effect): قد يؤدي الاعتماد المفرط على هذه الاختبارات إلى قيام المعلمين بتكليف تدرسيهم ليركز فقط على أنواع المعرفة والمهارات التي يمكن قياسها بسهولة بهذا النوع من الاختبارات (غالباً التذكر)، مما يعيق تحقيق الأهداف التعليمية الأوسع.

7. الارتباطات بمفاهيم ونظريات أخرى

تتصل الاختبارات القائمة على الاختيار ارتباطاً وثيقاً بالعديد من المفاهيم والنظريات الأساسية في مجالات القياس النفسي وعلم النفس التربوي. فهي تُعتبر جزءاً لا يتجزأ من منظومة أوسع للتقييم تهدف إلى فهم السلوك البشري والقدرات المعرفية. على سبيل المثال، يركز القياس النفسي على تطوير وتطبيق الاختبارات النفسية لقياس السمات، القدرات، والخصائص السلوكية، وتعد بنود الاختيار من متعدد مكوناً رئيسياً في العديد من هذه الاختبارات، مثل اختبارات الذكاء، الشخصية، والاستعدادات.

كما ترتبط هذه الاختبارات بنظريات التعلم، خصوصاً تلك التي تركز على السلوك والاستجابة، حيث توفر وسيلة موضوعية لقياس مخرجات التعلم. تُعد العلاقة بين جودة بنود الاختبار والموثوقية والصدق في القياس مفتاحية. فالصدق يشير إلى مدى قياس الاختبار لما صمم لقياسه فعلاً، بينما الموثوقية تتعلق باتساق النتائج. تصميم بنود اختيار من متعدد فعالة يساهم بشكل مباشر في تحقيق هذين المعيارين الأساسيين في أي أداة تقييم. علاوة على ذلك، تُستخدم هذه الاختبارات في إطار التقييم التشخيصي لتحديد نقاط القوة والضعف لدى المتعلمين، وفي التقييم التكويني والختامي لتقييم التقدم والتحصيل الكلي.

8. مبادئ صياغة بنود اختيار من متعدد فعالة

تعتمد قيمة ومصداقية المعلومات المستمدة من اختبارات الاختيار من متعدد بشكل مباشر وحاسم على جودة البنود المكونة لها. إن صياغة بنود فعالة تتجاوز مجرد كتابة سؤال وعدد من الإجابات المحتملة؛ إنها عملية تتطلب فهماً عميقاً للمحتوى، والمستوى المعرفي المستهدف، وخصائص المتعلمين، والمبادئ الأساسية في القياس التربوي. يتطلب الأمر مزيجاً من الخبرة الأكاديمية والمهارة الفنية لإنشاء أسئلة لا تقيس المعرفة فحسب، بل تُشجع أيضاً على التفكير النقدي وتُقدم رؤى قيمة حول تعلم الطلاب. فيما يلي مجموعة شاملة من الإرشادات الأساسية، المستمدة من الأبحاث الرائدة وأفضل الممارسات في مجال القياس، والتي يمكن أن تكون بمثابة دليل عملي لمساعدة القارئ على بناء بنود اختيار من متعدد ذات جودة عالية وفعالية قصوى (Haladyna et al., 2002; Downing & Haladyna, 2006; Gronlund & Waugh, 2009).

ربط كل بند بهدف تعليمي محدد وواضح: يجب أن يقيس كل سؤال مخرجاً تعليمياً مهماً ومحددًا مسبقاً من أهداف التعلم أو الكفاءات المستهدفة للمادة الدراسية أو البرنامج التدريبي. هذا يضمن أن الاختبار يظل مركزاً على الأولويات التعليمية ويزيد بشكل كبير من صدقه المحتوئي.

صياغة الجذع بوضوح ودقة متناهية: يجب أن يقدم الجذع مشكلة واضحة، محددة، ومفهومة بذاتها. يُفضل أن يكون الجذع ذا معنى قائم بذاته، بحيث يستطيع المفحوص فهم المطلوب بشكل كامل قبل الانتقال إلى قراءة البدائل. تجنب الصياغات الغامضة والنفي المزوج، واستخدم صيغة السؤال المباشر كلما أمكن.

التأكد من وجود إجابة صحيحة أو أفضل واحدة فقط: يُعد هذا المبدأ أساسياً لضمان موضوعية الاختبار. يجب أن تكون الإجابة المحددة كـمفتاح (Key) هي الصحيحة بشكل لا لبس فيه أو هي الأفضل بشكل واضح مقارنة بالبدائل الأخرى، وأن يكون هذا التحديد متفقاً عليه من قبل الخبراء في المجال.

صياغة مموهات (مشتتات) معقولة وجذابة: يجب أن تكون جميع المموهات معقولة (plausible) لتبدو كإجابات محتملة ومقبولة للمفحوصين الذين لم يتقنوا المادة بشكل كامل. يجب أن تستند المموهات الفعالة إلى الأخطاء الشائعة أو المفاهيم الخاطئة الشائعة التي يقع فيها المتعلمون.

الحفاظ على تجانس البدائل في الشكل والمحتوى: يجب أن تكون جميع البدائل (الإجابة الصحيحة والمموهات) متجانسة قدر الإمكان في محتواها، وشكلها، وطولها، وبنيتها النحوية. هذا يمنع المفحوص من استخدام تلميحات غير مقصودة لتخمين الإجابة الصحيحة.

تجنب التلميحات النحوية واللفظية غير المقصودة: تأكد من أن البدائل تتوافق نحويًا مع الجذع، ولا ينبغي أن يكون هناك ارتباط لفظي واضح أو تطابق في الكلمات الرئيسية بين كلمة في الجذع وكلمة في الإجابة الصحيحة لا يوجد مثله مع المموهات.

توزيع موقع الإجابة الصحيحة عشوائياً: يجب أن يظهر المفتاح (الإجابة الصحيحة) في كل موقع من مواقع البدائل بنسبة متساوية تقريباً على مدار الاختبار ككل، لتجنب الأنماط التي قد يستغلها المفحوصون للتخمين.

التركيز على قياس الفهم والتفكير، وليس فقط التذكر: كلما أمكن، صمم أسئلة تتطلب من المفحوصين تطبيق المفاهيم، تحليل البيانات، تفسير المعلومات، أو حل المشكلات، بدلاً من مجرد استدعاء الحقائق المعزولة.

مراجعة وتحكيم البنود بدقة: قبل استخدام البنود في اختبار فعلي، يجب مراجعتها بعناية فائقة من قبل الزملاء أو الخبراء في المحتوى والقياس، لضمان جودتها وصدق الاختبار وموثوقيته.

استخدام تحليل البنود بعد تطبيق الاختبار (Item Analysis): بعد تطبيق الاختبار، يجب استخدام البيانات الإحصائية لتحليل أداء كل بند على حدة (الصعوبة، القدرة التمييزية، فعالية المموهات)، واستخدام هذه المعلومات لتحديد البنود الضعيفة وتحسين بنك الأسئلة للاستخدام المستقبلي.

خاتمة: مستقبل التقييم القائم على الاختيار

لقد تناول هذا المقال الاختبارات القائمة على الاختيار كأداة أساسية وفاعلة في صندوق أدوات التقييم التربوي والمهني الحديث. تم التركيز بشكل خاص على أسئلة الاختيار من متعدد، نظراً لشيوعها، مرونتها، وقدرتها على قياس طيف واسع من المخرجات التعليمية. بدأنا رحلتنا بتحديد ماهية هذه الاختبارات وتمييزها عن غيرها من أساليب التقييم، ثم انتقلنا إلى تحليل دقيق لمكونات سؤال الاختيار من متعدد - الجذع، الإجابة الصحيحة، والمموهات - مبرزين الأهمية البالغة لكل عنصر في بناء سؤال فعال قادر على تحقيق أهدافه التقييمية.

استعرض المقال بعد ذلك الأنواع المتعددة لأسئلة الاختيار من متعدد، سواء من حيث تنوع صيغها البنائية كالنوع التقليدي،

والصح والخطأ المتعدد، والأسئلة المعتمدة على سياق، أو من حيث قدرتها على استهداف مستويات معرفية مختلفة تدرج من التذكر البسيط إلى الفهم، التطبيق، والتحليل. كما تم الإشارة إلى التحديات الكبيرة التي تواجه قياس المستويات العليا مثل التقويم والابتكار باستخدام هذا النوع من الأسئلة، مما يؤكد على ضرورة تصميمها بمهارة فائقة.

في المحصلة، تظل الاختبارات القائمة على الاختيار، وخاصة أسئلة الاختيار من متعدد، أداة تقويم قوية، فعالة، وقيمة لا غنى عنها في كثير من السياقات، شريطة أن يتم تصميمها واستخدامها بعناية، بحكمة، وفهم عميق لإمكاناتها وقيودها. إن الهدف ليس التخلي عن هذه الاختبارات، بل تحسين جودتها واستخدامها بشكل مسؤول كجزء لا يتجزأ من منظومة تقويم شاملة ومتنوعة تخدم أهداف التعلم والتطور بفعالية أكبر. ومع التقدم التكنولوجي المستمر، تظهر إمكانات جديدة لتطوير أشكال أكثر تطوراً وتفاعلية من بنود الاختبار، مثل الاختبارات التكيفية المحوسبة (Computerized Adaptive Testing - CAT) والبنود المبتكرة التي تدمج الوسائط المتعددة والمحاكاة، مما يبشر بمستقبل واعد لهذه الأداة التقويمية الهامة ويُعزز قدرتها على تقديم تقييمات أكثر دقة وواقعية (Parshall, Spray, Kalohn, & Davey, 2002).

المراجع

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). **A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of educational objectives**. Longman
- Bloom, B. S. (Ed.). (1956). **Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain**. David McKay
- Case, S. M., & Swanson, D. B. (2002). Constructing written test questions for the basic and clinical sciences (3rd ed., Rev.). National Board of Medical Examiners. <https://www.nbme.org/sites/default/files/2020-11/item-writing-manual.pdf>
- Cohen, R. J., & Swerdlik, M. E. (2018). **Psychological testing and assessment: An introduction to tests and measurement** (9th ed.). McGraw-Hill Education
- Downing, S. M. (2003). Constructing high-quality multiple-choice tests. In S. M. Downing & T. M. Haladyna (Eds.), **Handbook of test development** (pp. 281-300). Lawrence Erlbaum Associates
- Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & R. A. Yudkowsky (Eds.), **Assessment in health professions education** (pp. 3-25). Lawrence Erlbaum Associates
- Downing, S. M., & Haladyna, T. M. (Eds.). (2006). **Handbook of test development**. Lawrence Erlbaum Associates
- Ebel, R. L., & Frisbie, D. A. (1991). **Essentials of educational measurement** (5th ed.). Prentice Hall
- Frisbie, D. A. (1992). The multiple true-false format: A status review. **Educational Measurement: Issues and Practice**, 11(4), 21-26. <https://doi.org/10.1111/j.1745-3992.1992.tb00259.x>
- Gronlund, N. E., & Waugh, C. K. (2009). **Assessment of student achievement** (9th ed.). Pearson

Haladyna, T. M., & Downing, S. M. (1993). How many options is enough for a multiple-choice test item? **Educational and Psychological Measurement**, 53(4), 999-1010. <https://doi.org/10.1177/0013164493053004013>

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. **Applied Measurement in Education**, 15(3), 309-333. https://doi.org/10.1207/S15324818AME1503_5

.Nitko, A. J., & Brookhart, S. M. (2011). **Educational assessment of students** (6th ed.). Pearson

Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). **Practical considerations in computer-based testing**. Springer-Verlag

.Popham, W. J. (2017). **Classroom assessment: What teachers need to know** (8th ed.). Pearson

Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. **Practical Assessment, Research, and Evaluation**, 22(4). <https://doi.org/10.7275/ms65-6161>

Towns, M. H. (2014). Guide to developing high-quality, reliable, and valid multiple-choice assessments. **Journal of Chemical Education**, 91(9), 1426-1431. <https://doi.org/10.1021/ed500076x>