

F. L'habileté à répondre à des QCM

Certains étudiants même incompetents sur le fond, acquièrent une grande habileté à détecter la réponse correcte d'une QCM, et notamment à partir des erreurs de rédaction commises par l'auteur de la question.

1. Définition

Cette habileté est appelée, en anglais, *test wiseness*, que De Landsheere (1979) définit comme la *qualité de celui qui est aguerri aux tests*.

L'antonyme de *test wiseness* pose moins de problèmes de traduction : Thorndike (1971, p. 364) parle de *test naïveté* (en anglais).

Gibb (1964) définit la *test wiseness* comme « la capacité d'un étudiant à répondre correctement à un test fait de QCM qui contiennent des repères formels et, par conséquent, la capacité à obtenir des points sans connaître la matière testée ».

Nous verrons que l'étude expérimentale de cette habileté des étudiants est un des plus sûrs moyens de découvrir ou de valider des règles de rédaction des QCM. Il suffit de *transgresser* l'une de ces règles, en créant des questions volontairement biaisées, puis d'observer le comportement des étudiants.

2. Les méthodes d'étude de cette habileté

Deux techniques expérimentales ont été utilisées.

- a) Certains auteurs ont *introduit des erreurs de rédaction* dans des QCM bien rédigées (et qui avaient été éprouvées expérimentalement). Les résultats obtenus après l'introduction de l'erreur sont comparés aux résultats obtenus avec la forme initiale.
Board et Witney (1972) ont procédé de la sorte sur un test destiné à des étudiants universitaires ; Mc Morris, et al. (1972) ont fait de même sur un test destiné à des étudiants du secondaire supérieur.
- b) Le plus souvent, on propose à l'étudiant des *questions dont le contenu est fictif* (inventé de toutes pièces par l'auteur). Cette technique présente deux avantages majeurs :
 - 1° Il ne peut y avoir d'interférence entre les résultats et le degré de maîtrise de la matière chez les divers étudiants (puisque cette matière n'existe pas).

2° Le taux de choix de chaque solution devrait correspondre au hasard : par exemple, 25 % pour quatre solutions proposées. Toute *transgression* d'une règle de rédaction qui « *avantage* » une solution devrait se refléter directement dans le taux élevé de choix de celle-ci (plus de 25 % des fois). Dans les exemples qui suivent, nous marquerons d'un *astérisque* (*) la *solution* *avantagee*.

On trouvera dans la section G les principaux résultats obtenus au cours de diverses recherches utilisant des questions à contenu fictif. Les étudiants devaient soit choisir une solution, soit juger chaque solution proposée. Plusieurs de nos exemples seront empruntés aux trois études brièvement décrites ci-dessous.

Slakter, Koehler et Hampton (1970) ont transgressé quatre règles, chacune dans quatre QCM fictives. Les seize questions ont été mêlées à vingt-huit QCM normales et présentées à des étudiants de dix-sept ans. De plus, ces auteurs ont mesuré la *test wiseness* *avant et après un apprentissage spécifique* par enseignement programmé sur la façon de tirer parti de ces quatre erreurs.

Diamond et Evans (1972) ont transgressé cinq règles, chacune dans six QCM *fictives*. Le test de trente questions ainsi obtenu a été présenté à des élèves de douze ans.

Strang (1977) a introduit quatre QCM *fictives* parmi huit QCM normales, au niveau universitaire. Les étudiants étaient avertis que chaque solution devait être considérée en soi, car plusieurs pouvaient être correctes. Les étudiants pouvaient donc choisir plusieurs solutions par question.

Il est impossible de relater ici toutes les expériences faites dans ce domaine. Nous nous bornerons aux plus révélatrices d'entre elles à la fois par les résultats obtenus et par les moyens (souvent ingénieux) mis en œuvre pour les recueillir.

G. Résultats d'études expérimentales de l'habileté à répondre à des QCM

1. Transgression de la règle 8

Dans ces QCM l'article *an* (en anglais) et le pronom *il* (dans notre traduction française) fournit un indice facilitateur.

Exemple de question à contenu fictif (adapté de Diamond et Evans, 1972).

La dernière amitié de Harlan Stone est un

1. satire politique ?
- *2. récit autobiographique ?
3. histoire de science-fiction ?
4. courte biographie ?

Résultats

La solution avantageée a recueilli 35 % des choix (Diamond et Evans), et non les 25 % attendus. L'effet de l'article (*a* ou *an*) a aussi été mis en évidence par les études de Dunn et Goldstein (1959), et Mc Morris et al. (1972).

2. Transgression de la règle 14

Les solutions ne sont pas indépendantes.

a) Exemple de question fictive

La question ci-dessous est due à Diamond et Evans (1972). Si la solution 1 est correcte, les autres ne le sont pas.

La population de Frankton est inférieure à

- *1. 50 mille habitants
2. 60 mille habitants
3. 70 mille habitants
- *4. 80 mille habitants

b) Autre exemple de question fictive

La question ci-dessous est due à Slakter et al. (1972). Si la solution 2 est correcte, les solutions 1 et 4 le sont aussi. Seule la solution 3 est indépendante des autres : elle est donc avantagée.

Quand on ajoute des cristaux de Bestor à l'eau,

1. De la chaleur se dégage.
2. La température de la solution s'élève.
- *3. La solution devient bleue.
4. Le récipient devient plus chaud.

Résultats.

Au lieu des 25 % théoriquement attendus, la solution avantagée a reçu des pourcentages de choix égaux à

- 45 % dans l'expérience de Diamond et Evans (1972) ;
- 52 % avant l'apprentissage et 79 % après apprentissage dans l'expérience de Slakter et al. (1972).

3. Transgression de la règle 15

Un mot (ou une phrase) est répétée dans l'amorce et dans les solutions proposées.

Exemple de question fictive (due à Marshall et Hales, 1971)

Le Parti national augustin a son quartier général à

1. Camden, dans le New Jersey.
- *2. St Augustin, en Floride.
3. Palo Alto, en Californie.
4. Dallas, au Texas.

Résultats

- La solution avantagée a reçu
- 76 % des choix (Diamond et Evans, 1972)
 - 66 % des choix avant et 82 % après apprentissage (Slakter et al. 1972).

Des résultats semblables ont été obtenus par les études de Dunn et Goldstein (1959) et Mc Morris et al. (1972).

4. Transgression de la règle 16

Les solutions sont toutes absurdes, sauf une (qui, par le fait même est avantagée).

Exemple de question fictive (construite par nous afin d'illustrer le procédé)

Le vent du désert de Zubi est fait

1. De petits insectes.
2. De cristaux de cuivre.
- *3. De cristaux de glace.
4. D'aiguilles de pins.

Résultats

Dans une expérience de Slakter et al. utilisant ce genre de question, la solution avantagée a recueilli 84 % de choix *avant* et *après* apprentissage.

Le taux élevé de réponses correctes avant l'apprentissage et le peu d'effets de ce dernier laissent penser que la présence de distracteurs absurdes a rendu « évidentes » les questions fictives. Celles-ci ne sont donc peut être pas une bonne méthode pour étudier ce défaut de rédaction particulier.

5. Transgression de la règle 17

La solution correcte est plus longue que les distracteurs.

Exemple de question fictive (due à Marshall et Hales, 1971)

La loi Davis (XIXe siècle)

1. Fournit de l'argent aux écoles
2. Abolit une loi précédente
- *3. Interdit la fabrication, la vente, le transport ou l'utilisation de divers médicaments qui étaient alors utilisés dans des intentions illégales.
4. Augmenta le salaire des agents de l'Etat.

Résultat :

La solution avantagée a reçu 53 % des choix (Diamond et Evans, 1972). Des résultats semblables ont été obtenus par Chase (1964), Dunn et Goldstein (1959), Mc Morris et al. (1972).

L'expérience de Strang (1977) permettait d'étudier aussi l'effet de la longueur des solutions. Les solutions longues reçurent 57 % de réponses VRAI et les courtes 38 % de réponses VRAI, au lieu de 50 % théoriques.

6. Transgression de la règle 19

La présence de mots tels que *tous*, *toujours*, *jamais*, etc. dans certains distracteurs les rend peu attrayants.

Exemple de question à contenu fictif (due à Marshall et Hales, 1972)

Herman Klavemann est célèbre parce que

1. Il a développé toutes les échelles musicales utilisées en Occident.
2. Il a composé toutes les sonates pendant la période romantique.
3. Il a traduit tous les classiques russes en anglais.
- *4. Il a inventé l'épingle de sûreté.

Résultats

Dans l'expérience (due à Diamond et al.), la solution avantageée a recueilli 50 % des choix, au lieu des 25 % théoriquement attendus.

Dans l'expérience de Slakter et al., la solution avantageée a reçu 41 % des choix avant apprentissage (de « test wiseness ») et 75 % des choix après cet apprentissage.

7. Transgression de la règle 20

Le vocabulaire n'a pas le même degré de technicité dans toutes les solutions.

Exemple de question fictive (due à Strang 1977).

Dans la théorie de J. Havighurst

1. Les qualités *isofiliales* sont importantes.
2. Le jeu *acropratique* joue un grand rôle dans le développement de la personnalité morale de l'enfant.
- *3. Les peurs des enfants sont spécialement étudiées.
- *4. On considère la personnalité comme la dimension la moins bien comprise de l'ensemble de l'être humain.

Cette question permet d'étudier à la fois les effets de la règle 20 et ceux de la règle 17 (les solutions ne comportant pas de termes techniques sont « avantageées » ; les solutions 1 et 3 sont courtes, les solutions 2 et 4 sont longues).

On remarquera que même les termes techniques sont fictifs.

Dans l'expérience de Strang, l'étudiant doit juger chaque solution proposée.

Les solutions avantageées sont jugées vraies dans 60 % des cas (contre 50 % en théorie) alors que les « désavantageées » sont jugées vraies dans 35 % des cas (contre 50 % en théorie).

Dans la question ci-dessous, Strang a construit ces solutions pour présenter :

(en 1), aucun terme technique (donc tous termes courants).

(en 2), un terme technique courant (*).

(en 3), un terme technique relativement courant (*).

(en 4), un terme technique peu courant (*).

(*) que nous avons écrit en italique

Heinz Werner

1. a dressé le graphique du développement physique des enfants.
2. a mis au point une technique d'analyse des *chromosomes*.
3. a insisté sur l'importance de l'*hétéronomie*.
4. croyait que l'*aloplastie* était héréditaire.

Avant de construire les questions, l'auteur a établi expérimentalement à quelle catégorie (familier, vaguement familier, non familier) appartenait chaque terme technique pour les élèves considérés.

Des « juges » (étudiants du même âge) ont été invités à classer des termes techniques dans les trois catégories, à partir des définitions suivantes :

familier = je peux le définir grossièrement.

vaguement familier = je ne peux le définir, mais je l'ai déjà rencontré auparavant.

non familier = je ne me rappelle pas l'avoir rencontré.

Pour la question ci-dessus, les étudiants devaient de nouveau juger chaque solution proposée. Pour ces quatre types de solutions, les taux de jugement « VRAI » ont été, respectivement de 55 %, 41 %, 38 % et de 25 % (au lieu des 50 % théoriques).

8. L'ordre des solutions correctes dans un test

Les solutions correctes des questions successives forment une séquence systématique facilement repérable.

Rabinowitz (1970) a observé que les réponses successives aux diverses questions d'une épreuve ne sont pas entièrement aléatoires, même quand les étudiants sont entièrement ignorants. Par exemple, ils ont tendance à épuiser l'ensemble des possibilités avant de recommencer un « cycle ».

Ces observations sont confirmées par Wood (1977, p. 232) : il est très rare de rencontrer des sujets notant la même solution pour toutes les questions.

Wilbur (1970) a systématiquement fait varier la position de la solution correcte de 20 questions sur les 80 qu'en comptait un test. Dans la forme 1 du test, les 20 QCM avaient 1 comme solution correcte ; dans la forme 2, leur solution correcte était en deuxième position, et ainsi de suite. Les résultats ne montrent aucune influence de l'ordre, mais ils paraissent peu déterminants, car cet ordre systématique ne concernait qu'une question sur quatre, et par conséquent, n'a sans doute pas été perçu. Il est regrettable que l'auteur n'ait pas interrogé les étudiants à ce sujet.

9. Observations générales

D'autres observations, plus générales, ont été faites à l'occasion de ce type d'études.

1. Les infractions aux règles 12, 17 et 14 ont essentiellement pour résultat une augmentation de la facilité de la question. Leur influence sur la validité et la fidélité s'avère quasi nulle (Mc Morris, Brown et al., 1972), ce qui est surprenant. Board et Whitney observent, eux, une nette diminution de la validité et de la fidélité en cas d'infraction systématique.
2. La relation entre l'intelligence générale et l'habileté à tirer parti des défauts dans le respect des règles 12, 17, 14 semble faible (Diamond et Evans, 1972). On peut penser que même les moins intelligents sont capables de déceler des défauts aussi grossiers ! Certains étudiants sont plus sensibles à certains défauts de construction qu'à d'autres.
3. Les étudiants sont conscients de ces défauts (aux règles 12, 17, 14, 15 et 19) comme en témoignent leurs justifications a posteriori (Diamond et Evans, 1972, qui s'adressent à des étudiants de 12 ans ; Millman et al., 1965, qui s'adressent à des étudiants du secondaire).
4. Il est possible de « progresser » dans l'habileté à tirer parti de certains défauts, en s'entraînant au moyen de cours programmés appropriés. Néanmoins, ces progrès ne semblent pas se maintenir pour tous les types de défauts de rédaction (Slakter et al., 1970).
5. Selon Mc Morris et al. (1972), la méthode qui consiste à modifier des épreuves authentiques permettrait de mesurer l'habileté (test wiseness) chez des groupes d'étudiants, mais non chez des individus particuliers. D'autres recherches sur le même sujet ont été menées par Crehan, Koehler et Slakter (1974, Rowley (1974), Nilsson et Wedman (1978), ainsi que Jangsma et Pound (1977).