Consulter et interpréter le rapport d'analyses statistiques de Moodle

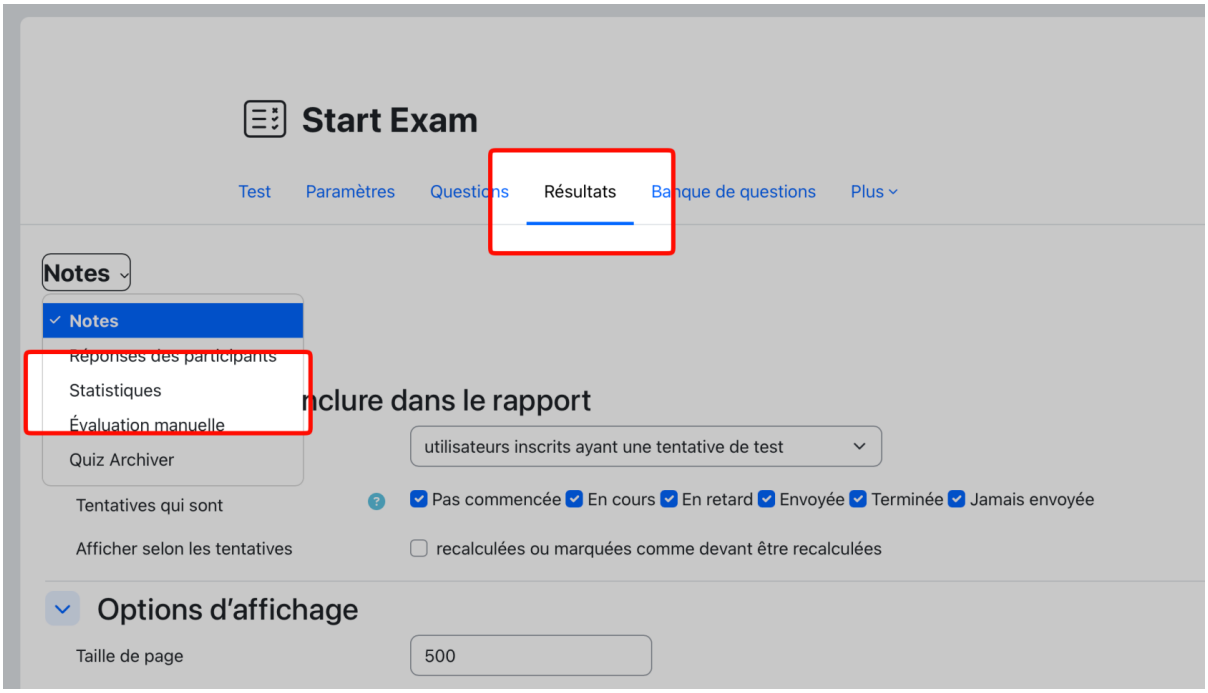
Introduction

> Ce nouveau rapport sera/est disponible dans MoodleUnil dès le 13 août 2026.

Lors de la réalisation d'un test – activité Test/Quiz ou QCM Papier) sur Moodle, une série d'indicateurs statistiques est automatiquement générée et mise à disposition. Ces indicateurs permettent d'évaluer la qualité de l'épreuve et de mettre en lumière certaines possibilités d'amélioration. Le présent document est un guide vous accompagnant dans la compréhension et l'interprétation de ces indices ainsi que de leur utilité, en vous proposant des clés de compréhension accessibles pour vous permettre de réaliser une analyse de la qualité de l'épreuve. Il est important de considérer les résultats de cette analyse comme des hypothèses à explorer, et non comme des verdicts définitifs ni de jugements sur le niveau des étudiant·es. L'utilisation de ces indicateurs permet de nourrir une démarche réflexive sur le dispositif d'évaluation que vous avez conçu. Pour toute question ou besoin d'un accompagnement plus approfondi, n'hésitez pas à contacter le CSE (cse@unil.ch).

1. Accès au rapport

Pour accéder au rapport d'analyses statistiques, accédez à l'activité correspondante (Test ou QCM Papier) sur Moodle, puis choisissez l'onglet « Résultats ». Choisissez ensuite « Statistiques » dans la liste déroulante



Le rapport se compose des éléments suivants :

1. Synthèse de l'analyse : affiche des informations globales telles que le taux de réussite, et met en évidence des questions un certain nombre de questions posant problème (le cas échéant).

2. Statistiques globales : affiche des informations quantitatives globales sur le test telles que date d'ouverture, score moyen, etc. Des infos-bulles apportent des informations sur le calcul et la signification des indicateurs statistiques.

3. Histogramme de la distribution des scores

4. Analyse de la structure du quiz : c'est là que sont affichés les indicateurs statistiques concernant chaque question. Des infos-bulles apportent à nouveau des informations sur le calcul et la signification des indicateurs statistiques.

2. Information du quiz

Les premiers indices statistiques sont calculés sur les scores totaux de l'examen. Les indices statistiques donnés dans le tableau « Information du quiz » sont détaillés dans cette section. Ils vous permettent de faire une première analyse globale de votre test qui pourra ensuite être approfondie (voir Analyse de la structure du quiz).

Score moyen

Score moyen (ou moyenne) est la somme des scores divisée par le nombre de tentatives. Il donne une idée de la tendance centrale des résultats. La moyenne peut être étirée (tirée vers le haut ou

le bas) par quelques valeurs extrêmes.

Pistes d'interprétation

- Une moyenne très basse peut indiquer un examen difficile.
- Au contraire, une moyenne très haute peut indiquer un examen facile.

Score médian

Score médian (ou médiane) est le score situé au milieu des résultats classés par ordre croissant. Le score médian représente donc la valeur pour laquelle 50% des étudiant·es ont un résultat supérieur et 50% un résultat inférieur. La médiane et la moyenne sont souvent comparées afin d'explorer s'il existe des valeurs extrêmes (scores élevés ou bas).

Pistes d'interprétation

- Si la moyenne est 14 et la médiane 10, cela signifie que quelques étudiant·es ont des scores élevés qui « tirent » la moyenne vers le haut. Quelques étudiant·es réussissent bien mieux l'examen, mais la majorité des étudiant·es ont un score plus bas.

Écart-type

L'écart-type mesure la dispersion des scores par rapport à la moyenne.

Pistes d'interprétation

- écart-type faible (proche de 0) : scores homogènes
- écart-type élevé (> 20% du score maximal) : variété des scores ; le test peut-être trop facile pour certain·es et trop difficile pour d'autres.

Fiabilité de l'examen - Alpha de Cronbach

Lors de la conception d'un test pour évaluer les apprentissages des étudiant·es, on souhaite que ce test mesure le plus fidèlement possible le niveau réel des connaissances des étudiant·es. Le coefficient α de Cronbach est un indicateur statistique qui permet d'évaluer la fiabilité interne du test, c'est-à-dire dans quelle mesure ses items (ses questions) forment un ensemble cohérent et donnent des résultats répétables dans des conditions comparables.

Pour illustrer ceci, prenons l'exemple d'un thermomètre. Un thermomètre fiable donne plusieurs fois la même température s'il est utilisé dans des conditions identiques. De la même manière, un test fiable devrait donner, à un·e étudiante dans des conditions similaires, des résultats proches. Dans la réalité, cette situation idéale n'est pas atteinte et il existe des fluctuations au niveau de résultats.

Qu'est-ce qui peut faire varier les résultats d'un test ?

- Le test lui-même : des questions mal formulées, ambiguës, qui ne mesurent pas toutes la même chose
- L'étudiant·e : la fatigue, le stress, la motivation, etc.
- Le contexte de passation : le bruit, l'heure , l'environnement. etc.

Le coefficient α se concentre sur la première source : les erreurs liées à la conception et à la cohérence des items. Il permet d'estimer quelle part de la variation des scores observés est attribuable à de vraies différences entre étudiant·es (ce qu'on veut mesurer), et quelle part est due à l'erreur de mesure (liée à la structure du test).

Par exemple, un coefficient α avec une valeur de 60% (α=0,60) signifie que 60% de la variation des scores peut être attribuée à des différences réelles de performance entre étudiant·es, et 40% peut être due à des fluctuations ou incohérences liées aux items du test. Autrement dit, le test a une fiabilité interne faible, ce qui peut inciter à retravailler les items pour améliorer cette cohérence.

Pistes d'interprétation

- Un indice > 70% est souvent considéré comme acceptable dans les sciences sociales.
- Un indice < 50% peut révéler une cohérence lacunaire du test : les items ne « fonctionnent » pas comme un ensemble cohérent.
- Un indice > 90% peut indiquer une grande corrélation entre certaines questions mais il peut aussi indiquer que les items sont trop similaires ou redondants (et donc ne mesurent pas assez de variabilité).

À garder en tête

- Le calcul de l'alpha de Cronbach suppose que tous les items mesurent un même construit(une seule dimension). Si le test évalue plusieurs dimensions différentes, la valeur obtenue peut être trompeuse.
- L'alpha de Cronbach mesure la cohérence interne, pas la validité : un test peut être très cohérent (valeur élevée) mais ne pas mesurer ce que vous avez voulu mesurer.
- L'augmentation du nombre d'items corrélés tend à augmenter α, mais cela ne remplace pas une bonne construction test-item (qualité des questions, pertinence, couverture des apprentissages). Inversement, avec un faible nombre de questions (en-dessous de 10), la valeur peut être basse sans que le problème vienne forcément de la cohérence du test.

Asymétrie de la distribution des scores

L'asymétrie (skewness en anglais) est un indicateur statistique qui décrit la forme de la distribution des scores, en particulier son degré de symétrie. Elle indique si les résultats sont répartis de manière équilibrée autour de la moyenne ou si la distribution "penche" vers un côté.

- Une asymétrie nulle signifie un équilibre approximatif.
- Une asymétrie positive ou négative indique la présence d'une queue plus longue d'un côté, souvent causée par quelques valeurs extrêmes (outliers).

Une distribution équilibrée n’est pas toujours nécessaire, mais elle est souvent souhaitable pour des tests d’apprentissage standardisés, car elle permet d’avoir une meilleure discrimination des niveaux de performance.

Pistes d’interprétation

- Asymétrie proche de 0 : distribution symétrique. Les scores sont distribués de manière relativement équilibrée autour de la moyenne. Cela ne signifie pas qu’ils sont uniformément répartis, mais qu’il n’y a pas de biais évident vers les faibles ou les hauts scores.
- Asymétrie positive : elle est caractérisée par une queue de distribution située vers la droite et signifie que la majorité des étudiant·es obtiennent des scores faibles ou modérés, tandis qu’une minorité atteint des performances très élevées. Ce type de distribution peut indiquer que le test est difficile, que les compétences évaluées ne sont maîtrisées que par les étudiant·es les plus préparé·es, ou encore qu’il existe quelques performances exceptionnellement bonnes qui tirent la distribution vers le haut.
- Asymétrie négative : Elle est marquée par une queue orientée vers la gauche et traduit une situation où la plupart des étudiant·es obtiennent de bons voire de très bons résultats, alors que quelques scores très faibles étirent la distribution. Cette configuration peut suggérer que le test est relativement facile, que les objectifs sont globalement maîtrisés par la majorité du groupe, ou qu’il existe quelques étudiant·es en difficulté ou insuffisamment préparé·es, dont les faibles performances créent l’asymétrie observée.

À garder en tête

Une asymétrie n’est pas forcément un problème :

- un examen très facile ou très difficile peut être volontaire ;
- certains apprentissages produisent naturellement des distributions asymétriques.

L’asymétrie peut être causée par des valeurs extrêmes plutôt que par un niveau général des étudiant·es.

L’asymétrie ne dit rien sur la qualité du test sur le plan de la validité ; c’est uniquement une information sur la forme de la distribution.

Kurtosis

Le kurtosis est un indicateur statistique qui décrit la forme verticale de la distribution des scores, c’est-à-dire son degré d’aplatissement ou de concentration autour de la moyenne. Il renseigne à la fois sur l’accumulation des résultats au centre et sur la fréquence des valeurs extrêmes. Une valeur proche de 3 (kurtosis non centrée) correspond à la distribution normale : ni trop pointue, ni trop étalée. Le kurtosis complète ainsi la moyenne et l’asymétrie pour comprendre la structure globale des performances.

Pistes d’interprétation

- Kurtosis = 3 : distribution normale et symétrique. La distribution est comparable à celle d'une loi normale : équilibre entre concentration centrale et présence modérée de valeurs extrêmes.
- Kurtosis > 3 : Une valeur supérieure à 3 indique que les scores sont fortement concentrés près de la moyenne, tout en présentant davantage de valeurs extrêmes (très hautes ou très basses) que prévu dans une distribution normale. Cette configuration peut refléter un examen globalement bien calibré mais contenant quelques questions très faciles ou très difficiles, qui génèrent des résultats atypiques.
- Kurtosis < 3 : Une valeur inférieure à 3 signale une distribution plus aplatie et dispersée. Les scores sont moins regroupés autour de la moyenne et davantage étalés sur l'ensemble de l'échelle. Cela peut indiquer un examen hétérogène, avec des questions dont la difficulté varie fortement, ou un test qui ne discrimine pas suffisamment les niveaux de performance.

À garder en tête

- Un kurtosis élevé (> 3) ne signifie pas que le test est "meilleur" : il peut simplement refléter la présence de plusieurs valeurs extrêmes.
- Un kurtosis faible (< 3) n'est pas forcément un problème : certains apprentissages produisent naturellement des distributions plus dispersées.
- Le kurtosis doit toujours être interprété avec l'asymétrie et le contexte du test : un examen peut être homogène mais très difficile, ou varié mais parfaitement valide.
- Une forte concentration autour de la moyenne peut indiquer une excellente calibration, mais aussi un manque de discrimination entre niveaux de compétence.
- À l'inverse, une très faible concentration peut signaler une variabilité excessive dans les items, mais peut également refléter une diversité réelle de niveaux dans la population étudiante.

Histogramme

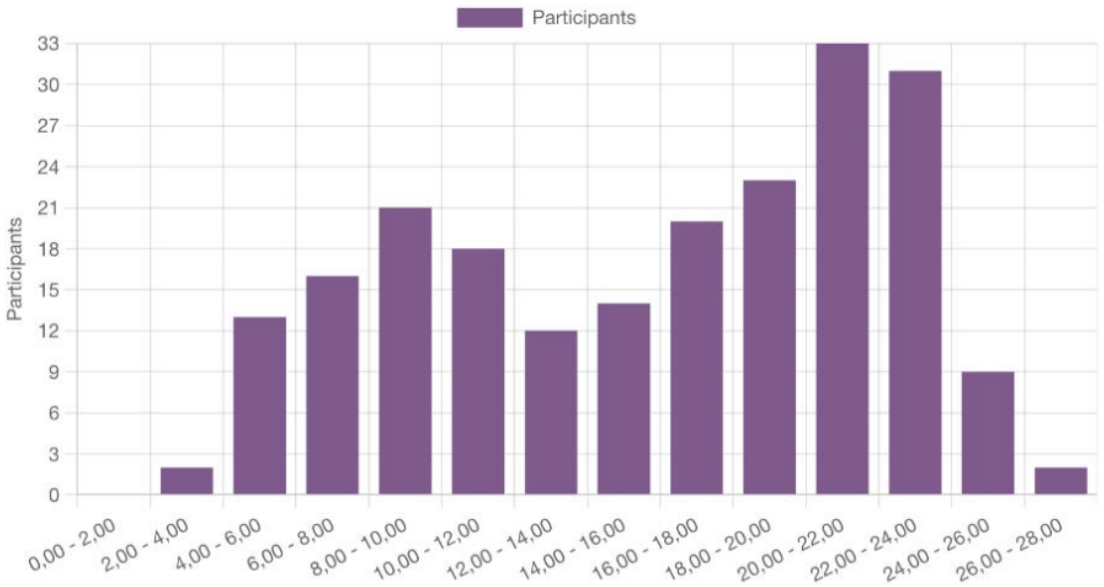
L'histogramme est un graphique qui présente la distribution des scores. Les observations sont réparties en différentes classes de même taille.

Pistes d'interprétation

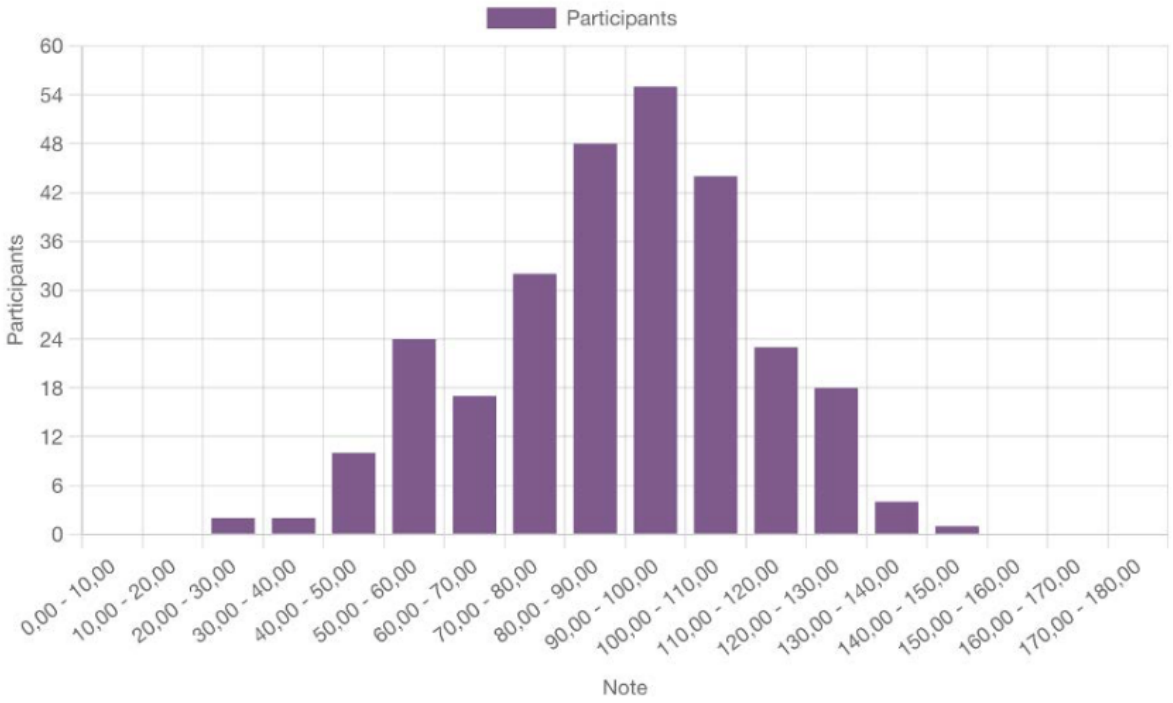
- L'histogramme permet d'avoir une représentation visuelle de la distribution des scores et de certaines tendances déjà renseignées par les indices présentés au-dessus

Exemples

Exemple 1 : écart-type 6,18 ; asymétrie -0,27 ; kurtosis -1,20



Exemple 2 : écart-type 22,42 ; asymétrie -0,31 ; kurtosis -0,22



3. Analyse de la structure du quiz

Dans cette seconde partie de l'analyse, les indices statistiques sont calculés pour chaque question. Ces indices permettent de mettre en évidence les bonnes questions et les questions qui semblent poser problème. Ces dernières sont celles qui demandent une analyse approfondie, notamment autour des enjeux pédagogiques.

Indice de difficulté

L’indice de difficulté correspond au pourcentage d’étudiant·es ayant répondu correctement à une question. C’est donc un indicateur direct du niveau de difficulté perçu de l’item. Plus ce pourcentage est élevé, plus la question est facile ; plus il est faible, plus elle est difficile. Cet indice est fondamental en psychométrie car il permet d’évaluer si les questions sont adaptées aux objectifs visés et si elles discriminent correctement les niveaux de maîtrise.

Interprétation des valeurs

- < 30% : question difficile
- > 70 % : question facile
- Entre 50% et 60% : zone idéale pour un examen qui vise à différencier les niveaux de compétences (plage recommandée en psychométrie)

Ces valeurs doivent toutefois être interprétées en fonction de la nature des objectifs d’apprentissage, du niveau du cours et de la logique globale de l’examen.

Pistes d’interprétation – Question difficile

Il est important de s’interroger sur la difficulté observée de cette question : est-elle attendue ou non ?

La question porte sur une compétence complexe, appliquée ou avancée, ou sur un savoir de haut niveau dans la taxonomie des apprentissages .

- Cela permet de distinguer les étudiant·es ayant une compréhension approfondie.
- Aucune analyse supplémentaire n’est nécessaire si cette difficulté était prévue.

Si la difficulté est inattendue, et notamment si la question porte sur un concept central du cours. Il est utile d’analyser les causes possibles du faible taux de réussite :

1. Revoir la formulation de la question et des propositions de réponse

- L’énoncé est-il clair et complet ?
- Tous les éléments nécessaires à la réponse sont-ils présents ?
- Les propositions de réponses sont-elles bien formulées ?

2. Examiner l’enseignement du concept

- Le concept évalué a-t-il été suffisamment abordé en cours ?

3. Les étudiant·es ont-ils eu l’occasion de s’entraîner sur des tâches similaires ?

Analyser les réponses des étudiant ·es :

- La bonne réponse est-elle malgré tout la plus fréquemment choisie ?
- Quelles propositions incorrectes attirent plus de 20% des étudiant·es ? Ces erreurs sont-elles compréhensibles (pièges logiques, confusion fréquente) ou plutôt liées à un défaut

de formulation ?

Pistes d'interprétation – Question facile

Quelques questions faciles (> 70 %) dans un examen ne posent généralement aucun problème, surtout si cela fait partie d'une stratégie d'évaluation, par exemple pour rassurer les étudiant·es ou évaluer des concepts « charnières ». Toutefois, un nombre important de questions faciles peut limiter la capacité du test à discriminer les niveaux de performance. Il est donc également intéressant de se questionner sur la facilité apparente d'une question. De plus, une question facile est régulièrement accompagnée d'une faible discrimination entre les étudiant·es (voir Indice de discrimination).

Si la facilité est attendue

La question teste un concept fondamental, indispensable à la suite du cours

- Il est positif que la majorité des étudiant·es y répondent correctement
- Pas d'analyse plus approfondie nécessaire

Si la facilité est excessive ou non souhaitée, il est pertinent de vérifier :

4. La formulation de l'énoncé et des réponses

- Contient des indices involontaires qui permettent de deviner la bonne réponse sans réelle compréhension?
- L'énoncé oriente-t-il inconsciemment vers la bonne option ?

5. La qualité des distracteurs

- Les distracteurs sont-ils crédibles ?
- Y a-t-il des Distracteurs Non-Fonctionnels (voir Distracteur non-fonctionnel) qui sont trop évidents ou peu crédibles ?

Indice de distraction

L'indice de distraction permet d'évaluer la qualité des distracteurs, c'est-à-dire les propositions de réponses incorrectes dans un QCM . Un distracteur efficace doit être crédible, attirer au moins une partie des étudiant·es qui ne maîtrisent pas la matière, et contribuer à différencier les niveaux de compréhension. À l'inverse, un distracteur trop facile à écarter perd son rôle pédagogique et affaiblit la capacité du QCM à mesurer correctement les apprentissages. La qualité des distracteurs est essentielle pour garantir la validité pédagogique d'un QCM.

Détection des Distracteurs Non Fonctionnels (DNF) et formule

Un distracteur est considéré comme non fonctionnel s'il est sélectionné par moins de 5% des étudiant·es.

- Cela signifie qu'il est trop évident, peu crédible, ou sans lien réel avec la tâche cognitive attendue.
- Un distracteur non fonctionnel n'apporte aucune information utile sur la maîtrise des étudiant·es.

L'indice de distraction est ensuite calculé sur la base des DNF pour chaque question selon :

$$1- \frac{(Nombre\ de\ DNF)}{(Nombre\ de\ propositions)}$$

- Plus l'indice est proche de 1, meilleure est la qualité des distracteurs.
- Plus il est faible, plus l'item contient de distracteurs non fonctionnels.

Pistes d'interprétation

- Indice de distraction < 40 % : proportion importante de distracteurs non fonctionnels → l'item pose problème.
- Indice ≥ 40 % : qualité globalement acceptable des distracteurs.

Piste d'analyse - Indice de distraction faible

Un indice de distraction faible indique que plusieurs distracteurs n'ont pas joué leur rôle. L'analyse doit alors déterminer si cela était attendu ou non.

Si la facilité est attendue

- L'item évalue un concept central, incontournable, que la plupart des étudiant·es doivent maîtriser.
- Les distracteurs deviennent mécaniquement moins attractifs car les étudiant·es maîtrisent réellement le contenu.

Dans ce cas, l'indice faible n'est pas problématique. Aucune analyse supplémentaire n'est nécessaire.

Si la facilité est excessive ou non souhaitée, l'analyse doit porter sur les points suivants :

1. La formulation de l'énoncé

- L'énoncé donne-t-il des indices involontaires ?
- Une formulation maladroite oriente-t-elle vers la bonne réponse ?

2. La qualité des distracteurs

- Les distracteurs sont-ils crédibles, cohérents avec le contenu ?
- Représentent-ils des erreurs plausibles, typiques des étudiant·es qui n'ont pas compris ?

- Y a-t-il des distracteurs trop absurdes, trop évidents ou sans lien avec le concept ?
- Leur absence de crédibilité peut expliquer le faible taux de sélection.

3. Analyse des choix des étudiant·es

- Certains distracteurs sont-ils totalement ignorés (DNF) ?
- Le bon choix est-il sélectionné quasi systématiquement, même par les étudiant·es ayant échoué ailleurs ?
- Cela peut révéler un item devenu trop simple par manque de distracteurs pertinents.

À garder en tête

- La qualité des distracteurs influence directement l'indice de discrimination et la capacité du test à distinguer les niveaux de maîtrise.
- Un distracteur non fonctionnel n'aide pas les étudiant·es, mais il nuit à la validité de l'épreuve en rendant la question artificiellement facile.
- Le rôle d'un distracteur n'est pas de "piéger", mais de rendre visibles les erreurs typiques liées à une mauvaise compréhension.
- L'objectif n'est pas d'augmenter artificiellement la difficulté mais de s'assurer que chaque distracteur corresponde à une fausse conception plausible.
- Un bon distracteur aide également l'enseignant à comprendre les difficultés conceptuelles des étudiant·es.

Indice de discrimination

L'indice de discrimination indique dans quelle mesure une question permet de différencier les étudiant·es selon leur niveau de performance globale à l'examen. Un bon item doit être davantage réussi par les étudiant·es qui maîtrisent la matière (ceux avec les meilleurs résultats à l'ensemble du test) et moins bien réussi par celles et ceux qui sont en difficulté. Cet indice reflète donc la capacité d'une question à participer réellement à la mesure des apprentissages.

Techniquement, il correspond à la différence entre : $\mathbf{P_x — P_y}$

Où :

- P_x: la proportion d'étudiant·es du groupe supérieur (27 % les mieux classés) qui réussissent la question ;
- P_y : la proportion d'étudiant·es du groupe inférieur (27 % avec les plus faibles scores) qui réussissent la même question.

Les valeurs possibles s'étendent de –100% à +100%

Piste d'interprétation - Discrimination faible (indice < 30%)

Une discrimination faible indique que la question ne parvient pas à distinguer les étudiant·es selon leur compréhension réelle. Les causes possibles sont :

1. Question trop facile

- Tout le monde réussit, quel que soit son niveau.

2. Question ambiguë ou mal formulée

- Les étudiant·es plus avancés peuvent être perturbés par une formulation floue ou des distracteurs peu crédibles.

3. Problème de qualité des distracteurs

- Distracteurs non fonctionnels ou trop évidents (voir Analyse des réponses).

4. Manque d'alignement

- L'item n'évalue pas réellement une compétence clé du cours.

Dans ce cas, il est recommandé de :

- revoir la formulation de l'énoncé ;
- vérifier l'indice de difficulté (question facile / difficile) ;
- analyser la qualité des distracteurs ;
- reformuler l'item avant un prochain usage.

Discrimination négative (indice < 0)

Une discrimination négative est anormale : elle révèle que les étudiant·es les moins performants réussissent mieux la question que les meilleur·es.

Les causes les plus fréquentes sont :

1. Erreur technique

- Mauvaise clé de réponse encodée dans MoodleExam (bonne réponse inversée).

Il s'agit de la première chose à vérifier.

2. Distracteurs trop convaincants

- Un distracteur peut être rédigé de manière si plausible qu'il attire particulièrement les étudiant·es avancés.

3. Énoncé ambigu ou trompeur

- Formulation mal alignée, piège involontaire, double négation, imprécision conceptuelle.

4. Désalignement pédagogique

- La question ne correspond pas réellement à ce qui a été enseigné ou évalué.

Dans ces cas, l'item doit être revérifié en profondeur. Si aucune erreur technique n'est détectée et que la question porte sur un concept-clé, il peut être pertinent de la retirer de l'examen pour l'établissement de la note finale.

À garder en tête

- L'indice de discrimination est un indicateur central pour évaluer la qualité pédagogique d'un QCM.
- Il doit toujours être interprété en lien avec l'indice de difficulté et l'indice de distraction.
- Un item avec une bonne discrimination contribue fortement à la validité du test : il mesure réellement des différences de compréhension.
- Une discrimination faible n'est pas forcément une erreur, mais un signal qu'il faut réexaminer la cohérence de l'item.
- Une discrimination négative est presque toujours un signe de problème (technique ou pédagogique).
- L'objectif n'est pas de rendre les questions difficiles, mais de s'assurer que chaque item permet de différencier les niveaux de maîtrise de manière fiable.

Fiabilité de l'examen si question supprimée – Alpha de Cronbach

Pour rappel, l'alpha de Cronbach (α) est un indicateur statistique qui mesure la fiabilité interne d'un test, c'est-à-dire la cohérence entre les différentes questions qui le composent (voir Alpha de Cronbach)

Ici, la valeur de l'indice est recalculée en retirant ladite question du test. Cette nouvelle valeur est ensuite comparée à la valeur pour l'ensemble du test (qui tient compte de toutes les questions).

Pistes d'interprétation

1. Si la suppression de la question augmente la valeur de l'alpha de Cronbach (par rapport à la valeur de base), cela signifie que cette question diminue la cohérence globale du test

- Plus la différence est grande, plus la question semble problématique
- Dans cette situation, plusieurs questions doivent se poser
- Quel concept cette question est-elle censée évaluer ?
- Le concept est-il bien enseigné dans le cours ?
- La formulation de la question est-elle claire et alignée avec les objectifs d'apprentissage ?

2. Une fois ces réflexions effectuées, vous pouvez alors :

- Reformuler l'énoncé et / ou les réponses pour un prochain examen
- Supprimer la question pour un prochain examen (Notamment si elle ne semble pas avoir de lien clair avec les objectifs d'apprentissage du cours)

4. Analyse des réponses

Cette troisième partie de l'analyse statistique concerne les propositions de réponse de chaque question.

Distracteur non-fonctionnel

En docimologie, les distracteurs (ou leurres) sont les propositions de réponse erronées. Un distracteur est considéré comme non fonctionnel s'il est sélectionné par moins de 5% des étudiant·es.

- Cela signifie qu'il est trop évident, peu crédible, ou sans lien réel avec la tâche cognitive attendue.
- Un distracteur non fonctionnel n'apporte aucune information utile sur la maîtrise des étudiant·es.

Si une question possède un ou plusieurs DNF, il est utile de se questionner par rapport aux points suivants :

1. La formulation de l'énoncé

- L'énoncé donne-t-il des indices involontaires ?
- Une formulation maladroite oriente-t-elle vers la bonne réponse ?

2. La qualité des distracteurs

- Les distracteurs sont-ils crédibles, cohérents avec le contenu ?
- Représentent-ils des erreurs plausibles, typiques des étudiant·es qui n'ont pas compris ?
- Y a-t-il des distracteurs trop absurdes, trop évidents ou sans lien avec le concept ?
  - Leur absence de crédibilité peut expliquer le faible taux de sélection.

Lors d'un prochain examen, envisagez de reformuler ou de remplacer les distracteurs non fonctionnels par des alternatives plus crédibles, en lien avec les erreurs typiques observées chez les étudiant·es.

Bonne réponse

On vérifie également que la bonne réponse ait la fréquence de sélection la plus élevée. Si ce n'est pas le cas, la question est mise en évidence en rouge .

Si ce cas est rencontré, il est utile de regarder les éléments suivants :

- Vérifier la bonne réponse et son encodage sur Moodle
- Pour les autres éléments, allez voir la partie Indice de difficulté

Cette situation est souvent le signe d'une question difficile.

Indice de distraction

L'indice de distraction est ensuite calculé sur la base des DNF pour chaque question selon :

$$1- \frac{(Nombre\ de\ DNF)}{(Nombre\ de\ propositions)}$$

- Plus l'indice est proche de 1, meilleure est la qualité des distracteurs.
- Plus il est faible, plus l'item contient de distracteurs non fonctionnels.

Pistes d'interprétation

- Indice de distraction < 40 % : proportion importante de distracteurs non fonctionnels → l'item pose problème.
- Indice ≥ 40 % : qualité globalement acceptable des distracteurs.

Piste d'analyse - Indice de distraction faible

Un indice de distraction faible indique que plusieurs distracteurs n'ont pas joué leur rôle. L'analyse doit alors déterminer si cela était attendu ou non.

Si la facilité est attendue :

- L'item évalue un concept central, incontournable, que la plupart des étudiant·es doivent maîtriser.
- Les distracteurs deviennent mécaniquement moins attractifs car les étudiant·es maîtrisent réellement le contenu.

Dans ce cas, l'indice faible n'est pas problématique. Aucune analyse supplémentaire n'est nécessaire.

Si la facilité est excessive ou non souhaitée, l'analyse doit porter sur les points suivants :

1. La formulation de l'énoncé

- L'énoncé donne-t-il des indices involontaires ?
- Une formulation maladroite oriente-t-elle vers la bonne réponse ?

2. La qualité des distracteurs

- Les distracteurs sont-ils crédibles, cohérents avec le contenu ?
- Représentent-ils des erreurs plausibles, typiques des étudiant·es qui n'ont pas compris ?
- Y a-t-il des distracteurs trop absurdes, trop évidents ou sans lien avec le concept ?
- Leur absence de crédibilité peut expliquer le faible taux de sélection.

3. Analyse des choix des étudiant·es

- Certains distracteurs sont-ils totalement ignorés (DNF) ?
- Le bon choix est-il sélectionné quasi systématiquement, même par les étudiant·es ayant échoué ailleurs ?
  - Cela peut révéler un item devenu trop simple par manque de distracteurs pertinents.

À garder en tête

- La qualité des distracteurs influence directement l'indice de discrimination et la capacité du test à distinguer les niveaux de maîtrise.
- Un distracteur non fonctionnel n'aide pas les étudiant·es, mais il nuit à la validité de l'épreuve en rendant la question artificiellement facile.
- Le rôle d'un distracteur n'est pas de "piéger", mais de rendre visibles les erreurs typiques liées à une mauvaise compréhension.
- L'objectif n'est pas d'augmenter artificiellement la difficulté mais de s'assurer que chaque distracteur corresponde à une fausse conception plausible.
- Un bon distracteur aide également l'enseignant à comprendre les difficultés conceptuelles des étudiant·es.

5. Pour aller plus loin

Caussarieu A., Jequier S., Rédiger des QCM efficaces – Un guide UNISCIEL (2022). UNISCIEL. Date, A. P., Borkar, A. S., Badwaik, R. T., Siddiqui, R. A., Shende, T. R., & Dashputra, A. V. (2019). Item analysis as tool to validate multiple choice question bank in pharmacology. *International Journal of Basic & Clinical Pharmacology*, *8*(9), 1999. https://doi.org/10.18203/2319-2003.ijbcp20194106

Hingorjo, M. R., & Jaleel, F. (s. d.). *Analysis of One-Best MCQs : The Difficulty Index, Discrimination Index and Distractor Efficiency.*

Malcourant, E., Raucent, B., & Wouters, P.. QCM or not QCM ? : *Processus de conception d'une évaluation par QCM* (2020). Presses universitaires de Louvain.

Rehman, A., Aslam, A., & Hassan, S. H. (2018). *ITEM ANALYSIS OF MULTIPLE CHOICE QUESTIONS. 38*(2). SMART, Guide méthodologique : Création d'une évaluation de qualité (2023). *Université de Liège*.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International *Journal of Medical Education, 2*, 53 55. https:/ doi.org/10.5116/ijme.4dfb.8dfd

Créé 2026-03-20 10:57:45 UTC par Nicolas Dunand
Mis à jour 2026-07-08 06:29:43 UTC par Nicolas Dunand