

Septembre 2024

La Certification Pix évolue : les algorithmes

Introduction

Pix a revu ses algorithmes pour la certification, l'obtention d'un score Pix et l'estimation du niveau par compétence. Il en résulte un nouveau test de certification composé d'un nombre constant de questions pour une durée maximum fixée.

Cette évolution est le fruit d'un travail de plusieurs années d'une équipe transverse interne impliquant toutes les compétences et métiers de Pix, notamment en collaboration avec Yvonnick Noël, psychométricien (Laboratoire de Psychologie, Cognition, Communication et Comportement, Université Rennes 2).

Cette nouvelle certification s'appuie sur 3 algorithmes :

- l'algorithme d'étalonnage des épreuves basé sur la théorie de la réponse à l'item[1][2][3],
- l'algorithme de choix des questions proposées en certification ou test adaptatif informatisé[4]
- l'algorithme de conversion de capacité en score et niveaux estimés par compétence.

Pix rejoint ainsi d'autres plateformes d'apprentissage et évaluation adaptatifs, comme Duolingo[5] ou TACIT.

Théorie de la réponse à l'item et étalonnage des épreuves

La théorie de la réponse à l'item (TRI)[1][2], est un pan des statistiques très fortement lié à l'éducation, où l'objectif principal est de mesurer les résultats individuels des apprenants. Elle traite de l'analyse statistique des réponses des personnes à un certain nombre d'items (ou questions). L'ensemble des compétences d'une personne n'étant pas observable directement dans un temps raisonnable, cette approche propose de les mesurer par l'estimation de la **capacité** des personnes à répondre à des questions étalonnées.

L'étalonnage basé sur la TRI consiste à définir 2 paramètres par question :

- sa **difficulté** relative, qui est corrélée à la proportion de personnes capables d'y répondre correctement,
- son **pouvoir discriminant**, qui représente la faculté de la question à être bien répondue par les personnes du niveau visé par la question, et non de façon aléatoire sur l'ensemble des répondants.

Pix a été créé en 2017, et compte presque 3 millions de nouveaux comptes par an sur les 4 dernières années, ce qui représente 1,5 milliard de réponses. Cela constitue une source de données d'une qualité et d'un volume uniques, sans équivalent dans le domaine des compétences numériques.

Une sélection des données a été effectuée suivant la mise en oeuvre de critères précis, pour éviter des biais et s'assurer d'un modèle représentatif de la population :

- des données récentes, datant de moins de 3 ans pour refléter l'état des connaissances numériques actuelles des candidats,
- élimination de réponses "non représentatives" et/ou susceptibles de fausser l'analyse, que ce soit par le temps de réponse, le faible nombre de participation d'un candidat, etc.

Tests adaptatifs informatisés

Dans un test adaptatif, les questions posées sont sélectionnées au cours du processus du test pour chaque personne testée. Ils sont conçus pour permettre à l'examineur d'avoir un test plus efficace que ce soit en nombre de questions à poser, mais aussi sur la précision de l'estimation de la capacité de la personne évaluée.

Le test adaptatif informatisé (TAI) nécessite 4 éléments pour son bon déroulement:

1. Une banque de questions étalonnées : le référentiel Pix (voir section précédente)
2. Une procédure pour choisir la prochaine question à poser
3. Une méthode de score qui sera détaillée dans la prochaine section
4. Une condition de fin du test.

Pour choisir la prochaine question, l'algorithme a besoin d'un critère de sélection. Pix utilise celui qui est donné en référence dans la littérature, l'Information de Fisher[7].

Ce critère permet de savoir quelle question sera la plus utile pour affiner l'estimation de la capacité de la personne testée.

Pour garantir la qualité et la pertinence du test d'évaluation des 16 compétences et proposer une évaluation accessible à tous les candidats, quel que soit leur niveau de maîtrise des compétences numériques, un ensemble de contraintes a été ajouté, notamment :

- Le TAI doit impérativement passer par l'intégralité des 16 compétences du référentiel ;
- Il ne peut passer qu'une seule fois sur une même thématique, évitant une redondance d'évaluation ;
- Une pondération de l'augmentation de la difficulté est appliquée permettant une progressivité dans la difficulté des épreuves proposées ;

Le développement et la mise en service d'un outil interne de simulation dédié à la Certification Pix a permis de vérifier, et permettra de vérifier sur le long terme, que les contraintes et paramètres ajoutés n'altèrent pas les résultats produits par le test adaptatif.

Calcul du score Pix et estimation du niveau par compétence

Score global

La conversion de la capacité en score pix se fait par une méthode en deux étapes pour assurer une flexibilité entre le référentiel étalonné et les points pix.

La première étape est la construction des mailles. Le score pix est composé de 8 mailles, correspondant aux 8 niveaux certifiés de Novice à Expert (64 à 1024 pix), de difficulté croissante. Une neuvième maille, pour les scores allant de 0 à 63 pix, regroupe les certifications pour lesquelles le premier niveau Novice 1 ne peut pas être délivré, lorsque les compétences mesurées sont insuffisantes.

Une cartographie des acquis du référentiel de compétences a été élaborée, à l'aide d'algorithmes d'aide à la décision, des données massives de résultats collectées par Pix et d'une analyse pédagogique fine par les équipes pédagogiques de Pix. Cette cartographie permet d'établir :

- les différentes mailles : composées d'un ensemble d'acquis estimés comme essentiels pour chaque niveau des compétences numériques (de Novice à Expert) ;
- les difficultés minimum et maximum de chaque maille.

La deuxième étape consiste à calculer un score de 0 à 1024, à partir de la position dans laquelle le candidat se trouve au sein de la maille à l'issue du test de

certification. Le score Pix donne donc les informations suivantes : le nombre de mailles maîtrisées et la progression dans celle en cours de maîtrise.

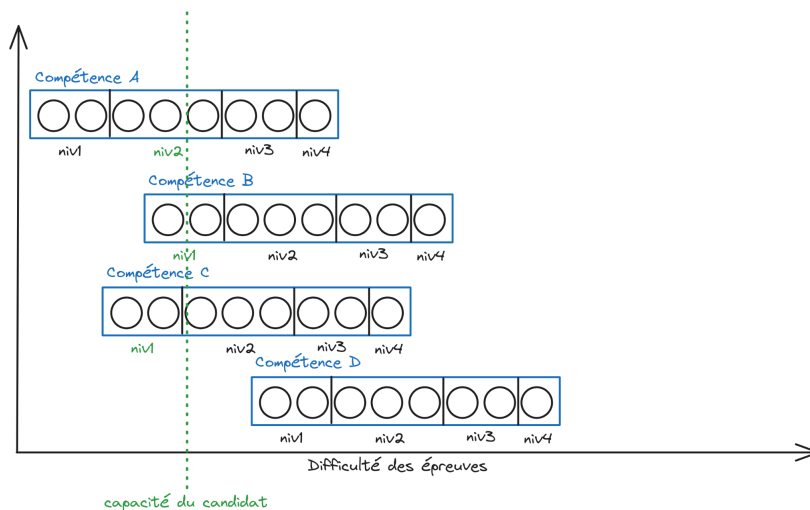
Niveau estimé des compétences

Avec plus de 3000 questions réparties dans 16 compétences et 7 niveaux par compétence, il est impossible d'être exhaustif dans l'évaluation de tous les niveaux de toutes les compétences de la personne candidate. Un compromis, outillé par la science, est d'estimer le niveau par compétence de la personne à partir de sa capacité estimée par le test adaptatif informatisé.

A l'instar des mailles, des seuils de difficulté par niveau et par compétence peuvent être définis. Ces seuils sont, là encore, fixés en mobilisant des algorithmes d'aide à la décision et des experts pédagogiques pour définir les acquis représentatifs d'un niveau.

Les compétences étant indépendantes les unes des autres, les seuils des niveaux sont différents d'une compétence à l'autre.

La capacité est transverse aux compétences, il est donc possible de relier le niveau d'une compétence à une plage de capacités. La définition des plages pour l'ensemble des paires niveau/compétence permet d'estimer le niveau par compétence des personnes ayant passé la certification.



Conclusion

La nouvelle certification de Pix repose sur 3 algorithmes qui sont issus de la rencontre entre :

- l'expertise pédagogique acquise par Pix en 6 ans d'existence ;
- les données massives collectées (1,5 milliard de réponses apportées par les utilisateurs de la plateforme Pix) ;
- les techniques d'apprentissage profond appliquées aux statistiques liées à l'éducation d'autre part.

Elle est une évolution du processus actuel qui prend en compte les retours d'expérience pour s'adapter aux besoins présents comme futurs.

Elle devient plus flexible pour s'adapter aux évolutions des compétences dans le domaine du numérique et donne les moyens à Pix de contribuer à un observatoire des compétences numériques.

Bibliographie

[1] Item Response Theory, R. Darrell Bock, Robert D. Gibbons, édition Wiley. Chapitre 3

[2] [Présentation Yvonnick Noel](#)

[3] Rocher, Thierry. « Mesure des compétences. Méthodes psychométriques utilisées dans le cadre des évaluations des élèves », MENS-DEPP, 2015.

[4] Item Response Theory, R. Darrell Bock, Robert D. Gibbons, édition Wiley. Chapitre 8

[5] Machine Learning–Driven Language Assessment, Burr Settles and Geoffrey T. LaFlair, Masato Hagiwara 2020

[7] [Information de Fisher](#)

Glossaire

Capacité : métrique obtenue à la suite d'un test adaptatif informatisé donnant une indication sur les connaissances d'une personne.

Difficulté : valeur numérique qui permet de comparer les questions entre elles en les classant par proportion de réussite parmi les répondants. Plus cette valeur est élevée, plus la question nécessite une expertise pour y répondre.

Pouvoir discriminant : Pour une même difficulté, permet de comparer la capacité des bons répondants aux questions. Plus cette valeur est élevée, plus la capacité des bons répondants est élevée.