10 Des évaluations éducatives par le jeu

Jack Buckley, Laura Colosimo, Rebecca Kantar, Marty McCall, et Erica Snow
Imbellus, Inc., États-Unis

Ce chapitre examine comment les progrès récents de la technologie numérique pourraient conduire à une nouvelle génération d'évaluations éducatives par le jeu. Les systèmes d'éducation disposeraient alors d'évaluations capables de tester des compétences plus complexes que les tests standardisés classiques. Après avoir souligné certains des avantages des évaluations par le jeu par rapport aux autres tests, ce chapitre aborde la manière dont ces tests sont construits, comment ils fonctionnent, mais aussi certaines de leurs limites. Si les jeux présentent un grand potentiel pour améliorer la qualité des tests et étendre l'évaluation à des compétences complexes à l'avenir, ils viendront probablement compléter les tests classiques, qui ont aussi leurs avantages. Trois exemples d'évaluations par le jeu qui intègrent des technologies avancées illustrent cette perspective

Introduction

Les récents développements technologiques tels que la réalité virtuelle/augmentée, l'interface utilisateur numérique et la conception d'expériences, l'apprentissage automatique/l'intelligence artificielle ainsi que l'exploration des données éducatives ont permis d'améliorer les environnements numériques simulés et d'accélérer les progrès en matière de qualité et de conception des simulations numériques et des jeux vidéo. Nous disposons aujourd'hui d'applications d'« apprentissage en ligne » à utiliser dans le cadre scolaire et en dehors de celui-ci (p. ex., les laboratoires virtuels ou les outils d'apprentissage médical en ligne avec simulation), mais les récents progrès technologiques ont également ouvert la voie à une nouvelle génération d'évaluations standardisées. Ces évaluations par le jeu permettent d'évaluer un plus large éventail de compétences (p. ex., la créativité, la collaboration ou les compétences socio-émotionnelles), ainsi que de mieux mesurer certains aspects de la « pensée » des répondants, y compris dans des domaines traditionnels comme les sciences et les mathématiques. En outre, l'utilisation d'environnements simulés permet d'évaluer les connaissances et les compétences dans des contextes qui correspondent à la mise en pratique de ces compétences dans la « vraie vie ».

Même si cette nouvelle génération d'évaluations semble promise à un bel avenir, elle n'en comporte pas moins ses propres limites. Par exemple, ces évaluations sont plus coûteuses et plus difficiles à développer que les tests standardisés traditionnels basés sur une simple succession de questions précises ou de petites tâches. Néanmoins, certaines évaluations par le jeu standardisées ont déjà été développées avec succès et feront probablement partie des évaluations de demain. Le chapitre est organisé de la manière suivante : nous soutenons tout d'abord que ces nouvelles évaluations répondent à de nombreuses critiques envers les évaluations classiques et qu'elles ont le potentiel de mieux correspondre aux besoins de l'enseignement et l'apprentissage en classe. Nous expliquons

ensuite comment ces évaluations fonctionnent, quel type de technologie et de données elles utilisent et nous soulignons les défis liés à leur élaboration. Nous fournissons enfin quelques exemples de ces évaluations innovantes standardisées, avant d'envisager le rôle qu'elles pourraient avoir à l'avenir et l'infrastructure nécessaire pour les proposer à grande échelle.

Pourquoi proposer des évaluations éducatives par le jeu ou par simulation ?

L'utilisation d'évaluations standardisées dans l'éducation - de plus en plus associées à des normes bien définies par rapport au contenu académique - est loin d'être une idée nouvelle, puisqu'elle remonte à une quarantaine d'années dans certains pays à revenu élevé et à au moins 20 ans au niveau international (Braun et Kanjee, 2006[1]). Plus récemment, des responsables dans le domaine de la politique de l'éducation, de l'enseignement et de l'apprentissage, et de la théorie cognitive se sont réunis pour demander davantage de cohérence entre l'enseignement, les programmes de cours et les évaluations, ainsi qu'un système d'évaluation complet qui oriente les décisions, de l'appareil législatif aux établissements scolaires (Gong, 2010[2]). Dans le cadre de cette initiative, on observe un intérêt croissant envers les nouvelles technologies et approches d'évaluation, notamment les évaluations immersives, par le jeu ou par simulation (DiCerbo, 2014[3] ; Shaffer et al., 2009[4] ; Shute, 2011[5]). Comme nous l'expliquons ci-dessous, ces nouvelles approches tirent parti de la diffusion croissante des technologies éducatives dans les établissements scolaires, ainsi que des avancées en matière de psychométrie, de conception d'évaluations informatisées, d'exploration de données éducatives et d'apprentissage automatique/intelligence artificielle qui bénéficient aux concepteurs de tests.

Dans le domaine de l'éducation, les évaluations standardisées classiques ont longtemps été dominées par un modèle centré sur des séries de questions précises (également appelées « items ») conçues pour couvrir le contenu d'un cadre d'évaluation en abordant les parties du domaine à mesurer (Mislevy et al., 2012[6]). En revanche, les évaluations par le jeu cherchent à réduire le fossé entre les évaluations classiques et les activités d'apprentissage plus stimulantes grâce à l'utilisation de jeux et de simulations conçus pour mesurer des constructs dans un environnement qui maximise la « fluidité » et récompense les élèves pour avoir démontré leurs processus cognitifs dans des situations plus stimulantes et authentiques, et pas seulement leur capacité à mémoriser des éléments clés (Shute et al., 2009[7]).

Alors que les évaluations éducatives classiques sont conçues pour répondre généralement à des normes de qualité technique telles que la validité (l'évaluation mesure-t-elle ce qu'elle est censée mesurer ?), la *fiabilité* (le fait-elle de manière cohérente et avec un minimum d'erreur ?) et *l'équité* (l'évaluation est-elle culturellement sensible, accessible et exempte de tout biais à l'égard de participants aux tests ?), ces évaluations ont été critiquées sous plusieurs angles. Nous examinons brièvement les différentes critiques émises à l'encontre des évaluations standardisées traditionnelles (p. ex., Sanders et Horn, 1995[8]) et la manière dont l'évaluation basée sur le jeu peut les améliorer :

- nécessité d'appliquer la théorie psychologique moderne aux évaluations ;
- manque de cohérence entre les évaluations et les programmes de cours (Duncan et Hmelo-Silver, 2009[9]) ;
- manque d'intégration des évaluations à des fins différentes, notamment formatives, partielles et sommatives (Perie, Marion et Gong, 2009[10]) ;
- incapacité des évaluations classiques à mesurer certains constructs importants et de plus en plus pertinents pour les politiques éducatives (Darling-Hammond, 2006[11]), et ;
- baisse de l'implication et de la motivation des étudiants (Nichols et Dawson, 2012[12]).

Application de la théorie psychologique moderne aux évaluations

L'ouvrage fondamental Knowing *What Students Know* (National Research Council, 2001[13]) a introduit la théorie cognitive dans le domaine de l'évaluation en utilisant un cadre accessible aux enseignants et aux décideurs. Il préconisait l'examen des fonctions mentales impliquées dans la compréhension profonde, des concepts difficiles à évaluer avec le type de questions courtes et déconnectées, fréquemment utilisées dans les tests standardisés (Darling-Hammond et al., 2013[14]). Cet ouvrage demandait la mise en place de nouveaux types de tâches (ou d'items fréquemment utilisés en classe, mais pas dans les tests normalisés) qui mobilisaient des compétences plus poussées, y compris des dissertations, portfolios, projets pédagogiques et observations des performances en classe. Les jeux et les simulations ont gagné en importance en raison de leur capacité à générer des éléments probants en matière de compréhension approfondie et de processus cognitifs. L'interprétation des données en

continu provenant du jeu ou de l'interaction avec une interface utilisateur numérique soigneusement conçue permet aux chercheurs d'évaluer comment les apprenants s'y prennent pour résoudre des problèmes et peut conduire à un retour d'information mieux ciblé (Chung, 2014[15]). Par exemple, les normes scolaires modernes en termes de contenu scientifique exigent de plus en plus que les élèves apprennent et démontrent des faits et pratiques scientifiques (c'est-à-dire qu'ils pensent et raisonnent comme un scientifique). Dans ce sens, les évaluations par le jeu permettent aux concepteurs des tests de créer des scénarios et des simulations permettant d'observer le raisonnement et les processus suivis par les élèves grâce à leurs interactions complexes avec les éléments du jeu ou de la simulation.

Nécessité de mieux adapter les évaluations aux programmes de cours et à l'enseignement

Conformément à l'objectif d'améliorer l'éducation, les programmes de cours ont évolué. Ils intègrent désormais les théories de l'apprentissage et des approches conceptuelles centrées sur les preuves qui offrent aux étudiants des expériences de base et des exemples pratiques (Mislevy et al., 2012[6] ; Arieli-Attali et al., 2019[16]). Cependant, les évaluations standardisées classiques sont restées relativement figées, ne fournissant que des informations limitées aux enseignants et aux apprenants, et accentuant le fossé entre ce qui est appris (contenu du programme de cours) et ce qui est effectivement testé (contenu des évaluations) (Martone et Sireci, 2009[17]). Les chercheurs et les décideurs continuent de réclamer de nouveaux cadres d'évaluation qui comprennent les théories de l'apprentissage et des compétences fondamentales transférables cohérentes avec l'activité en classe (National Research Council, 2012[18] ; Darling-Hammond et al., 2013[14] ; Conley, 2018[19]), ce qui a suscité un intérêt accru envers le développement de jeux, de simulations et de systèmes de tutorat intelligents conçus autour de processus d'apprentissage ou d'unités pédagogiques spécifiques.

Améliorer la cohérence des évaluations

Les évaluations sont en général classées en fonction de leur objectif : comment les scores sont-ils utilisés et interprétés ? Les tests sommatifs sont administrés à la fin d'une leçon ou d'un chapitre de cours pour évaluer ce qui a été appris. Parmi les exemples d'applications de l'évaluation sommative, citons les épreuves annuelles à grande échelle et les examens d'entrée à l'université, mais aussi les enquêtes comme PISA, TIMSS et diverses évaluations nationales (Oranje et al., 2019[20]). Les évaluations sommatives peuvent représenter des enjeux élevés pour les apprenants (examens d'entrée à l'université ou de fin d'études). En règle générale, elles représentent souvent un faible enjeu pour les élèves, mais un enjeu plus élevé pour les autres acteurs du système éducatif. Les évaluations intermédiaires sont organisées pendant la période d'enseignement pour évaluer les progrès en fonction des objectifs sommatifs et suggérer des changements dans l'approche pédagogique. Les évaluations formatives sont également faites pendant l'enseignement, mais sont étroitement liées à un enseignement spécifique et à la performance individuelle. Contrairement aux évaluations intermédiaires qui peuvent être regroupées à différents niveaux d'enseignement et sont liées à des objectifs sommatifs plus larges, les évaluations formatives sont ajustées aux besoins individuels et à la stratégie d'enseignement à court terme (Shepard, Penuel et Pellegrino, 2018[21]). Chacun de ces niveaux d'évaluation éducative a des objectifs différents et nécessite souvent des modèles de mesure et des méthodes de validation appropriés (Ferrara et al., 2017[22]).

La combinaison de tous ces différents types d'évaluation peut être source de confusion pour les professionnels de l'éducation et les parents et se fait souvent au détriment du temps d'instruction des élèves. C'est dans ce contexte que l'on cherche désormais à rationaliser ce système un peu confus et fragmenté. Aux États-Unis, par exemple, alors que les examens fondés sur la théorie et pertinents sur le plan de l'enseignement sont devenus plus importants, de nombreux appels ont été lancés en faveur d'une plus grande cohérence en matière d'évaluation (Gong, 2010[2] ; Marion et al., 2019[23]). En d'autres termes, les décideurs politiques et les professionnels de l'éducation souhaitent de plus en plus que toutes les évaluations auxquelles les élèves participent tout au long de l'année scolaire fonctionnent ensemble comme un système unique et cohérent.

Alors que les jeux et les simulations utilisés dans les évaluations ont le plus souvent ciblé le niveau formatif, les progrès récents dans le développement et la notation ont rendu possible leur utilisation dans les tests sommatifs des systèmes nationaux et des enquêtes internationales (Verger, Parcerisa et Fontdevila, 2019[24] ; Klieme, 2020[25]). Dans un système plus cohérent, une évaluation immersive par le jeu pourrait être utilisée de différentes manières. Sur le plan formatif, les évaluations par le jeu pourraient permettre de fournir un retour continu et des suggestions personnalisées tout au long de l'apprentissage. À titre de mesure intermédiaire, les élèves peuvent être évalués

dans des conditions de simulation plus standardisées afin de mesurer leurs progrès par rapport aux objectifs sommatifs. En mode sommatif, on pourrait présenter aux élèves un scénario d'évaluation par le jeu, un scénario inhabituel, mais en rapport avec le sujet. Ils seraient invités à le résoudre sans soutien formatif, ce qui permettrait de mieux comprendre ce que les élèves ont appris et sont capables de faire. Encadré 10.1 présente un exemple d'évaluation des compétences professionnelles en Allemagne.

Encadré 10.1 Simulations pour la formation et l'évaluation des compétences professionnelles en Allemagne

Avec les projets ASCOT+, le ministère fédéral allemand de l'Éducation et de la Recherche soutient le développement de la formation et de l'évaluation numériques des compétences professionnelles dans différents domaines (mécatronique automobile, résolution de problèmes pour les systèmes techniques, résolution de problèmes commerciaux, compétences interprofessionnelles et socio-émotionnelles en soins infirmiers). En plus des unités de formation numérique qui utilisent des vidéos et des simulations, le projet développe des évaluations en guise d'examens pour certifier les compétences des apprentis. Par exemple, dans le domaine des professions commerciales, un créateur de tâches d'évaluation axées sur les compétences est en cours de développement pour permettre aux évaluateurs de concevoir des examens qui certifient les compétences des apprenants, permettant la mise en place d'un examen basé sur les compétences, plutôt que sur les connaissances. Les évaluateurs pourront s'appuyer sur une banque de tâches numériques d'évaluation pouvant être légèrement modifiées ou combinées à des fins de personnalisation. Elle sera lancée en 2022 (et officiellement reconnue en Allemagne). Dans le domaine de la mécatronique automobile, des tâches d'examen sont également développées pour tester les compétences des stagiaires dans un environnement simulé et aussi pour renforcer leurs compétences.

Source: Bundesministerium für Bildung und Forschung (s.d.[26]).

Mesure de différents constructs - compétences « difficiles à mesurer »

Une autre critique formulée à l'encontre des évaluations classiques est qu'elles sont inefficaces pour mesurer les connaissances, les compétences et les aptitudes au-delà de contenus très simples, dans des domaines circonscrits (Madaus et Russell, 2010[27]). Par exemple, si un test standardisé classique peut être un moyen valide, fiable, équitable et efficace pour évaluer des connaissances en algèbre, il peut ne pas être approprié pour mesurer des constructs tels que la pensée créative ou la résolution collaborative de problèmes. Cette critique s'avère particulièrement pertinente dans le domaine de l'éducation, et ce pour deux raisons. Premièrement, les programmes pédagogiques modernes du monde entier sont de plus en plus multidimensionnels ; ils incluent à la fois des compétences transversales et des contenus académiques plus traditionnels. Par exemple, les normes scientifiques de prochaine génération aux États-Unis (www.nextgenscience.org/) comprennent non seulement des concepts de base dans une matière spécifique, mais aussi des idées transversales en sciences et en ingénierie. Deuxièmement, les décideurs internationaux prennent de plus en plus conscience de l'importance des « compétences du XXIe siècle » ou des compétences associées à un « apprentissage plus approfondi », telles que l'esprit critique, la communication, la collaboration et la créativité (Trilling et Fadel, 2009[28] ; Vincent-Lancrin et al., 2019[29] ; Fadel, Bialik et Trilling, 2015[30]). L'utilisation de jeux ou de simulations est un moyen très prometteur d'évaluer ces constructs complexes, soit dans le cadre d'un programme de cours révisé, soit comme un complément au contenu couvert par les tests standardisés habituels (Stecher et Hamilton, 2014[31] ; Seelow, 2019[32]).

Mesurer les constructs différemment - interaction et implication

Il est établi que la plupart des personnes qui passent des tests n'apprécient pas trop les évaluations classiques (Nichols et Dawson, 2012[12] ; Madaus et Russell, 2010[27]). Outre les avantages déjà évoqués ci-dessus, un des attraits des évaluations basées par le jeu réside dans leur capacité à fournir une mesure valide et fiable de constructs complexes tout en proposant une forme d'implication et d'immersion propre aux jeux vidéo modernes.

Même si de nombreux éléments probants viennent conforter cet avantage dans un large éventail d'évaluations par le jeu (Hamari et al., 2016[33]), il faut néanmoins garder à l'esprit qu'il y a une différence majeure entre les jeux pratiqués pour le plaisir et les jeux visant à mesurer certains constructs (en particulier, mais sans s'y limiter, ceux utilisés dans des contextes à enjeux élevés). En outre, les évaluations par le jeu doivent répondre à des critères scientifiques plus stricts en matière de validité, de fiabilité et d'équité, ce qui signifie que la transférabilité de l'implication et de l'immersion peut être quelque peu limitée ou du moins de nature différente (Oranje et al., 2019[20]). Pour faire bref, les évaluations par le jeu pourraient ne pas être aussi amusantes que les « vrais » jeux.

Nous allons maintenant examiner de plus près les caractéristiques de telles évaluations et envisager brièvement leur mode de conception.

Comment élaborer des évaluations par le jeu ?

Concevoir à partir du terrain

L'utilisation des jeux et de leurs caractéristiques comme moyen d'accroître l'implication des répondants et de saisir des constructs difficiles à mesurer n'est pas une idée nouvelle (Cordova et Lepper, 1996[34]). Cependant, les connaissances dans le domaine de l'évaluation, ainsi que la compréhension de la meilleure façon de mettre en œuvre ce type d'évaluation et d'utiliser au mieux les données qui en découlent continuent d'évoluer. Il existe de nombreuses façons d'incorporer des jeux et des caractéristiques propres à ceux-ci dans un système ou une évaluation. Ainsi, l'élaboration d'une évaluation par le jeu requiert une réflexion préalable sur les types exacts de caractéristiques que l'on veut mettre en avant et leur impact potentiel sur l'apprenant et la collecte de données (Shute et Ventura, 2013[35]).

Le concepteur de l'évaluation doit déterminer, a priori, ce que le jeu tente de mesurer exactement et comment chaque élément fournit des éléments probants qui viennent renforcer l'objectif initial. Cela veut dire qu'il faut « scénariser » les mesures dignes d'intérêt, déterminer les éléments qui viendront étayer ces mesures et la quantité exacte de ces éléments. Comme Mislevy (2018[36]) le soutient, « *la pire façon de concevoir une évaluation par le jeu ou une simulation est de concevoir ce qui ressemble à un super jeu ou une super simulation, de recueillir quelques observations pour obtenir toutes les informations que les performances au jeu fournissent, puis de confier les données à un psychométricien ou à un scientifique des données pour qu'ils leur assignent une note.* » Même si les analyses exploratoires a posteriori présentent des avantages, elles ne devraient pas être l'élément moteur qui détermine la façon dont on note l'évaluation. Avant que les concepteurs de l'évaluation ne commencent à élaborer les spécifications du jeu, ils doivent d'abord définir ce qu'ils veulent mesurer et comment cela sera fait. Cela inclut la quantification des éléments à vérifier et les échelles qui seront utilisées.

Cet important travail ne peut pas être effectué a posteriori, car il en résulterait souvent des performances psychométriques médiocres ou un manque d'interprétabilité. Par exemple, si l'adaptation d'un jeu existant en vue de son utilisation comme évaluation peut, à première vue, sembler générer une grande quantité de données pour chaque personne testée, il arrive souvent que ces données produisent des items ou des possibilités de mesure qui ne sont pas en phase avec le domaine à évaluer, qui présentent une forte intercorrélation (rendant bon nombre d'entre eux inutiles) ou qui ne sont pas au bon niveau de difficulté (c'est-à-dire trop faciles ou trop difficiles pour la population cible). Par conséquent, le processus de conception des items devrait avoir lieu plus tôt dans le projet, car la conception d'une évaluation par le jeu exige beaucoup de réflexion et de rigueur, et les égarements peuvent s'avérer particulièrement coûteux.

Toutefois, cela ne veut pas dire que l'analyse des données réelles des participants au test dans le cadre du processus d'élaboration des évaluations par le jeu n'est pas importante. Non seulement les concepteurs doivent-ils effectuer les analyses psychométriques empiriques traditionnelles nécessaires à la création d'évaluations valides et fiables, mais ils doivent également tirer parti de la richesse des données supplémentaires générées par les évaluations par le jeu pour appliquer de nouvelles méthodes issues de domaines tels que l'apprentissage automatique afin d'extraire des informations plus exploitables sur les aptitudes des personnes testées ou d'autres constructs lorsque cela est possible, voir (Gobert, Baker et Wixon, 2015[37]).

Jeux pour évaluation contre « ludification »

Nous établissons ici une distinction importante entre la conception de jeux ou de simulations explicitement à des fins de mesure et la « ludification » ou l'ajout d'éléments ludiques à des tâches ou à des activités existantes afin

d'accroître l'implication ou la motivation (Deterding et al., 2011[38]). Un exemple de ludification serait l'ajout d'un tableau de classement, de badges, d'avatars personnalisés ou de barres de progression à une activité en classe et, bien que cela puisse être utile pour améliorer l'implication ou la motivation des élèves, ce n'est pas le type d'évaluation par le jeu dont nous parlons ici. Il est important de noter, cependant, que la distinction entre ces dernières et la ludification est un peu moins nette dans l'évaluation de l'apprentissage socio-émotionnel ou des compétences « non cognitives » dans l'éducation et le monde professionnel. Néanmoins, il est toujours possible d'évaluer ces autres types de compétences et de dispositions via des jeux spécifiquement développés à cet effet (Yang et al., 2019[39]) et non simplement via l'ajout d'éléments ludiques à des tests traditionnels.

La télémétrie et la question de la « furtivité »

Les évaluations par le jeu ou la simulation permettent de recueillir une multitude de données qui sont souvent négligées ou impossibles à saisir via les tests classiques - parfois de manière « furtive » ou à l'insu de la personne qui passe le test (Shute et Ventura, 2013[35]). Il s'agit notamment de modèles de choix, de comportements de recherche, du temps passé sur une tâche et, dans certains cas, de mouvements oculaires ou d'autres informations biométriques. Ces riches sources de données peuvent servir à illustrer les processus cognitifs dans lesquels les élèves se lancent lorsqu'ils effectuent une tâche (Sabourin et al., 2011[40] ; Snow et al., 2015[41]), plutôt que de se concentrer uniquement sur le résultat final de leur performance. Cependant, afin de recueillir et de quantifier ces informations, les développeurs des évaluations par le jeu doivent définir avec soin les données que le système recueille, souvent appelées « télémétrie ». Ce processus implique qu'il faut répertorier chaque action qu'un utilisateur peut effectuer pendant la phase de conception et lui attribuer une valeur ou un nom dans l'infrastructure de données. Pour ce faire, on utilise le plus souvent des cadres de collecte de données ou de mesure tels que la conception centrée sur les preuves [Evidence-Centred Design] (Mislevy et al., 2012[6]). La mise en correspondance de la télémétrie avec les objectifs de mesure nécessite un effort concerté entre les concepteurs, les ingénieurs logiciels et les scientifiques de la mesure. Comme pour toute évaluation, les parties prenantes doivent avoir confiance dans ce qui est mesuré et dans la façon dont c'est fait. Si nous voulons utiliser la télémétrie dans les évaluations éducatives - en particulier dans les applications sommatives et à enjeu élevé - nous devons être très clairs sur les actions qui sont enregistrées, leur interprétation et la manière dont elles doivent être stockées et quantifiées.

Est-ce réellement difficile à mettre en place ? Quels sont les coûts par rapport aux approches plus traditionnelles ?

Notre expérience nous donne à croire que la mise au point d'évaluations par le jeu qui soient valides, fiables et équitables est considérablement plus complexe et difficile que le développement de tests standardisés classiques. Pour obtenir un résultat probant, il faut une équipe interdisciplinaire dotée d'un large éventail de compétences, notamment des concepteurs de jeux, des ingénieurs logiciels ayant idéalement une formation en jeu et des spécialistes des sciences cognitives, ainsi que des concepteurs de tests, des experts en contenu, des chercheurs en éducation et, enfin, des psychométriciens. Le développement d'une évaluation par le jeu est donc relativement coûteux et n'est pas nécessairement le meilleur moyen de mesurer des constructs de base. À titre d'exemple, même si les avantages des évaluations par le jeu ont conduit des programmes d'évaluation, comme le PISA (OCDE) et le National Assessment of Educational Progress (NAEP) des États-Unis, à ajouter des composantes de jeu ou de simulation dans le matériel d'enquête, ils l'ont fait de façon limitée, en raison du coût, en combinant nouveautés et items plus traditionnels (Bergner et von Davier, 2018[42]).

Un autre défi à relever est de rendre les évaluations par le jeu utilisables et accessibles aux élèves souffrant de déficiences. De nombreux progrès ont été réalisés à cet égard ces dernières décennies. Si l'on veut élargir des cadres d'évaluation comme la Conception universelle de l'apprentissage (CUA) (Rose, 2000[43]) aux évaluations par le jeu, il faudra une conception minutieuse, des tests approfondis et, dans certains cas, inventer de nouvelles approches et de nouvelles technologies telles que les dispositifs haptiques qui mettent en œuvre des interfaces utilisateur tactiles et permettent l'évaluation des étudiants malvoyants (Darrah, 2013[44]).

Nouvelles méthodes psychométriques et défis

Les évaluations par le jeu font appel à un éventail plus large de compétences techniques, mais requièrent également des innovations technologiques et de nouvelles approches en termes de mesure. Par exemple, les psychométriciens ont proposé de nouveaux modèles de mesure reflétant la complexité des

tâches (Mislevy et al., 2000[45] ; Bradshaw, 2016[46] ; de la Torre et Douglas, 2004[47]). Ceux-ci et d'autres en cours de développement visent à mieux prendre en compte les théories de la cognition et de l'apprentissage et à saisir les structures latentes complexes des aptitudes des apprenants. Ils fournissent des modèles de mesure adaptés aux nouveaux flux de données générés par les jeux et les simulations.

Les évaluations éducatives par le jeu soulèvent également de nouvelles préoccupations en matière de justice et d'équité. Concrètement, l'accès aux ordinateurs, dans le cadre familial ou scolaire, n'est pas universel. La maîtrise de la logique des jeux vidéo ou de l'interface utilisateur n'est pas uniforme non plus. Ces éléments pourraient creuser davantage les écarts de résultats existants ou en créer de nouveaux. Le développement responsable des évaluations par le jeu doit tenir compte de ces différences et minimiser le fonctionnement différentiel des items (FDI - c'est-à-dire le fait que les items ne se comportent pas comme prévu pour des personnes ayant la même capacité, mais des expériences différentes) en fonction des sous-groupes habituels (sexe, ethnicité, statut linguistique), mais aussi de nouveaux autres sous-groupes potentiels, comme l'expérience des jeux vidéo (voir l'Encadré 10.2). Un élément fondamental qui réduit le risque de fonctionnement différentiel des items dans les évaluations par le jeu est la conception de tutoriels efficaces dans chaque jeu ou simulation, qui enseignent rapidement les mécanismes du jeu aux personnes qui sont peut-être moins familières avec l'interface utilisateur.

Encadré 10.2 L'importance du choix de la technologie pour l'égalité entre les sexes : aperçu d'une expérience au Chili

Les résultats d'une expérience menée dans un établissement public de Santiago suggèrent que les différences de genre dans l'apprentissage des jeux éducatifs peuvent dépendre de la plateforme technologique utilisée (Echeverría et al., 2012[48]). Ce constat peut également s'appliquer aux évaluations par le jeu. Dans le cadre de cette expérience, des élèves de 11e année ont joué à First Colony, un jeu éducatif qui demande aux élèves d'appliquer des concepts d'électrostatique. Les joueurs se mettent dans la peau d'astronautes envoyés en mission pour rapporter un cristal précieux. Le cristal étant fragile, les astronautes ne peuvent le déplacer qu'en utilisant la force électrique. Dans la version du jeu mise en œuvre sur une plateforme équipée de plusieurs souris d'ordinateur, les élèves jouent par groupes de trois, chaque élève contrôlant une souris. À l'aide de celle-ci, les élèves peuvent déplacer leur astronaute, modifier la valeur et la polarité de leur charge et activer leur charge pour interagir avec le cristal. Dans la version en réalité augmentée, les élèves peuvent effectuer les mêmes actions à l'aide d'une tablette. La salle de classe se fond ici dans l'univers du jeu : chaque pupitre est recouvert d'un ensemble de marqueurs qui permettent au système de réalité augmentée de placer des objets virtuels sur les bureaux. Grâce à la webcam située en haut de l'écran, le système détermine l'emplacement de l'astronaute de chaque élève en détectant la position relative de chacun par rapport aux marqueurs en papier. Alors qu'aucune différence de performance entre les sexes n'a été observée lorsque les élèves ont joué à l'aide de la plateforme à souris multiples, les garçons ont obtenu de meilleurs résultats que les filles lorsqu'ils ont joué au même jeu à l'aide d'une plateforme de réalité augmentée, avec une différence statistiquement significative.

Les résultats de l'expérience ont révélé des différences de performance statistiquement significatives entre les garçons et les filles après l'utilisation de la plateforme de réalité augmentée. Étant donné qu'il n'y avait pas d'écart entre les sexes après l'utilisation de la version du jeu à souris multiples, cela donne à penser que le choix de la plateforme peut créer un écart entre les sexes dans l'apprentissage qui n'est pas lié au jeu proprement dit. Il est donc vraisemblable que l'utilisation de la technologie de la réalité augmentée désavantage les filles. Les professionnels de l'éducation doivent donc choisir avec soin la technologie utilisée dans les évaluations par le jeu.

Les perspectives de l'IA et de l'apprentissage automatique

Au-delà des innovations en matière de psychométrie, les évaluations par le jeu et par simulation offrent également de nouvelles possibilités d'innovation technique en lien avec les récents développements de l'apprentissage

automatique et de l'intelligence artificielle (Ciolacu et al., 2018[49]). Par exemple, dans les applications à enjeu élevé nécessitant une évaluation à plusieurs variantes pour garantir la sécurité, les algorithmes d'intelligence artificielle (IA) à forte intensité de calcul permettent de calibrer la difficulté des variantes d'un jeu pour garantir l'équité pour tous les participants aux tests. En d'autres termes, on peut faire appel à l'IA pour « simuler » toutes les variantes proposées d'une évaluation afin d'augmenter la probabilité qu'elles soient toutes comparables en termes de difficulté, avant de passer à des tests pilotes coûteux et longs avec des participants humains.

De manière plus générale, un mécanisme d'IA similaire, ainsi que l'application de techniques d'apprentissage automatique sur des jeux de données concernant la performance au jeu, peuvent être utilisés pour obtenir des informations pertinentes à partir des journaux de données télémétriques (c'est-à-dire les données recueillies au cours du processus d'évaluation par le jeu/la simulation). En d'autres termes, une partie essentielle du développement des évaluations par le jeu devrait inclure une étape où les scores sont affinés et améliorés par le biais d'une analyse exploratoire des données et d'une fouille des données éducatives, à mesure que de plus grandes quantités de données sur les participants aux tests deviennent disponibles. Même si la fouille de données ne doit pas remplacer le processus de conception décrit ci-dessus, les expériences suggèrent que l'itération assistée par ordinateur peut améliorer la fiabilité et l'efficacité des évaluations par le jeu en augmentant la quantité d'informations utiles sur la performance des participants (Mislevy et al., 2014[50]).

Quelques exemples d'évaluations éducatives par le jeu

SimCityEDU : Pollution Challenge (GlassLab)

SimCityEDU : Pollution Challenge est une évaluation par le jeu publiée en 2014 par GlassLab. Il s'agit d'une initiative de développement collaboratif financée par les fondations John D. et Catherine T. MacArthur et Bill et Melinda Gates. Conceptuellement, il s'inspire de la série de jeux SimCity et place le joueur dans le rôle du maire d'une ville virtuelle, chargé de concilier développement économique et protection de l'environnement sur une série de quatre niveaux de complexité croissante. SimCityEDU a été conçu comme une évaluation formative de la résolution de problèmes, de la pensée et de la causalité systémiques s'adressant aux élèves du niveau du premier cycle de l'enseignement secondaire.

Le contenu est expressément conforme aux normes du cadre d'évaluation de l'apprentissage au XXIe siècle et du Conseil pour l'enseignement et s'inspire également des normes scientifiques de nouvelle génération et des normes mathématiques de base. Tout comme dans le jeu initial, les tâches de résolution de problèmes sont très intéressantes et en grande partie de nature spatiale et économique. GlassLab a également consacré des ressources considérables à la résolution de problèmes tels que la « tutorisation » et le traitement de la télémétrie afin de créer des items d'évaluation utiles, ainsi qu'à la mise au point de nouveaux modèles psychométriques pour favoriser les inférences et la production de rapports (Mislevy et al., 2014[49] ; Mislevy, 2018[36]).

Graphique 10.1 SimCityEDU : Pollution Challenge (GlassLab)

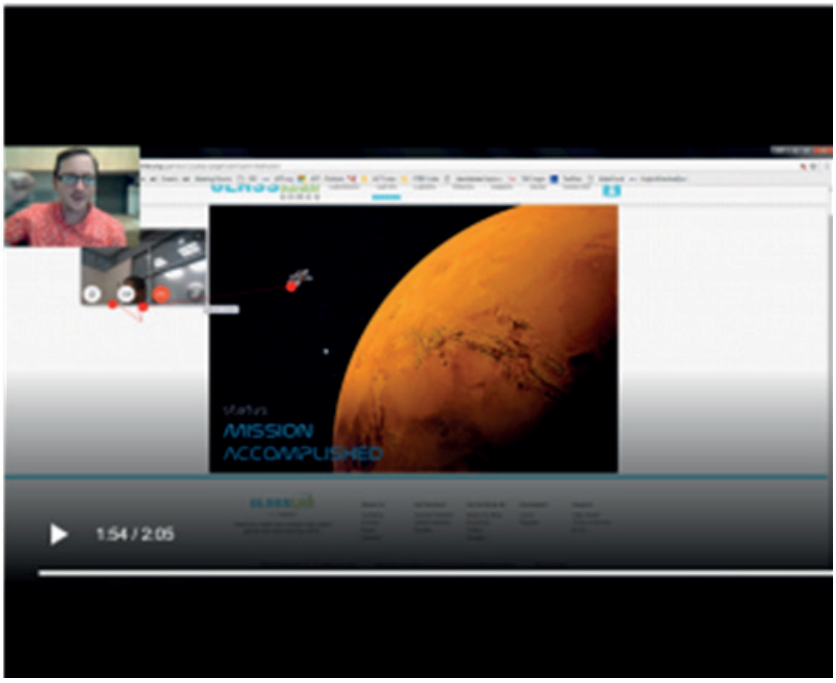
Remarque : Lancé en 2013 par la société aujourd'hui disparue GlassLab, SimCityEDU : Pollution Challenge est une version adaptée de la célèbre franchise de jeux vidéo SimCity et est destinée aux élèves du premier cycle de l'enseignement secondaire. Les élèves utilisent une version modifiée de l'interface de SimCity pour résoudre divers problèmes urbains. Les données télémétriques sont traitées par des modèles psychométriques sophistiqués.
Source: Games for change (s.d.[51]); Glasslab (s.d.[52]).

Crisis in Space (ACTNext)

Dans Crisis in Space, ACTNext a développé une version pilote d'une évaluation par le jeu afin évaluer la résolution collaborative de problèmes et les compétences socio-émotionnelles des élèves du premier cycle du secondaire. Dans ce jeu, deux élèves doivent travailler ensemble pour résoudre une série de problèmes dans une station spatiale : l'un joue le rôle d'un astronaute dans la station et l'autre contrôle la mission au sol. En demandant aux élèves de collaborer réellement dans un jeu coopératif, Crisis in Space offre une expérience authentique et attrayante tout en améliorant les tentatives précédentes de mesurer la collaboration par le biais d'une interaction élève-agent (chatbot).

Crisis in Space, qui a remporté le prix de l'innovation aux 2020 e-Assessment Awards, est particulièrement remarquable parce qu'il utilise un large éventail de données, y compris la télémétrie générée par l'interface utilisateur, les enregistrements des conversations des élèves et les données de suivi oculaire des joueurs. ACTNext a également mis en œuvre une technologie d'apprentissage automatique avancée, telle que le traitement du langage naturel (TLN), pour traiter ces données et évaluer les moments de collaboration (Chopade et al., 2019[53]).

Graphique 10.2 Crisis in Space (ACTNext)



Remarque : Crisis in Space est une évaluation pilote par le jeu développée par ACT, Inc. dans le cadre d'un programme de recherche et de développement de l'évaluation collaborative de la résolution de problèmes mené par la branche de recherche, ACTNext. Dans le scénario, deux joueurs travaillent ensemble pour dépanner une station spatiale. Les technologies utilisées pour évaluer le comportement des joueurs comprennent le suivi oculaire et le traitement du langage naturel.
Source: https://actnext.org/collaboration-assessment-online-games/ ; https://actnext.org/research-and-projects/cps-x-crisis-in-space/ (reproduit avec autorisation).

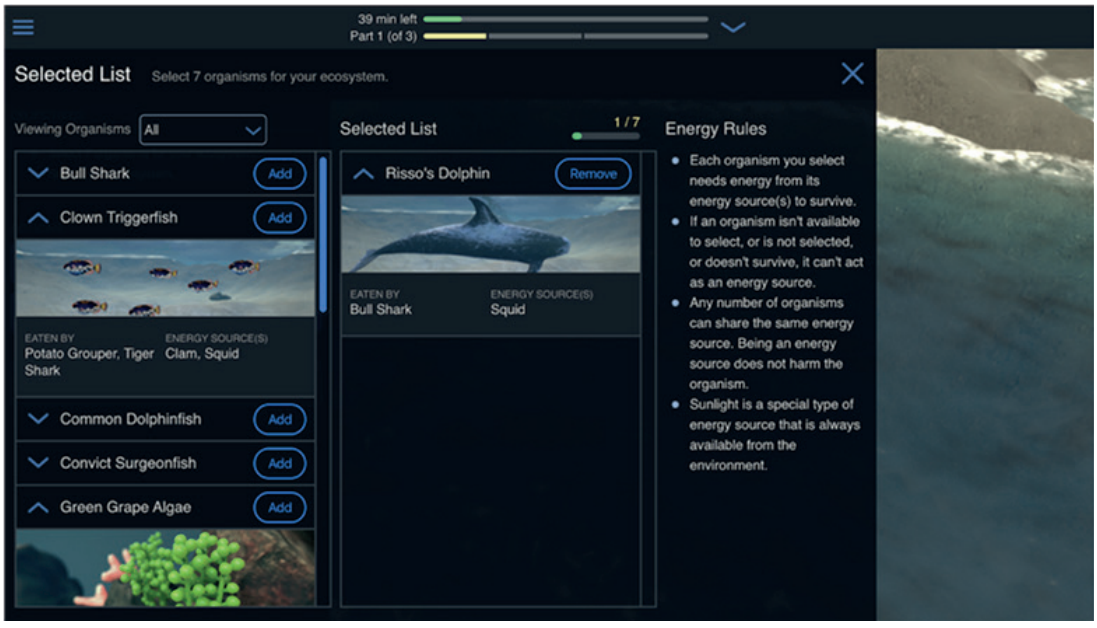
PEEP – Project Education Ecosystem Placement (Imbellus)

Le troisième exemple d'évaluation par le jeu, *Project Education Ecosystem Placement* (PEEP), est également en phase pilote et vise à mesurer la résolution de problèmes chez les élèves de l'enseignement secondaire grâce à un jeu dans lequel les participants doivent construire un réseau alimentaire ou un écosystème viable et le placer dans un environnement naturel où il peut se développer. *PEEP*, financé par la fondation de la famille Walton, est la version adaptée d'une évaluation par le jeu, initialement conçue pour recruter du personnel et actuellement utilisée par le cabinet de conseil international McKinsey pour sélectionner de nouveaux analystes de gestion.

La version éducative a été adaptée pour prendre en compte des sujets propres aux sciences de la vie et mieux correspondre à l'apprentissage des élèves. Le *PEEP* est conçu pour être éventuellement utilisé dans le cadre

d'évaluations sommatives à enjeu élevé et permet la création de nombreuses versions parallèles afin d'améliorer la sécurité des tests. Dans cette perspective, le *PEEP* utilise un algorithme pour créer des solutions d'écosystèmes viables de difficulté approximativement équivalente à partir d'une grande bibliothèque d'organismes. Le *PEEP* peut également être administré en tant que tâche d'évaluation « adaptative par étapes », où l'on présente aux candidats une série de problèmes à résoudre dont la difficulté varie de manière algorithmique en fonction des performances antérieures.

Graphique 10.3 PEEP - Project Education Ecosystem Placement (Imbellus)



Remarque : Conçu pour être utilisé dans le cadre d'une évaluation plus large de la résolution de problèmes, le PEEP demande aux élèves de construire un écosystème viable et de le placer dans un environnement naturel où il peut se développer. L'évaluation est une adaptation d'un autre produit développé par Imbellus pour être utilisé dans le recrutement. La fondation de la famille Walton a soutenu le projet, en partenariat avec *Summit Public Schools* et d'autres systèmes scolaires. La télémétrie des élèves est utilisée pour évaluer les processus mobilisés et offrir la possibilité de noter les résultats selon la théorie de la réponse à l'item.
Source: Avec l'aimable autorisation d'Imbellus.

Quelles sont les perspectives à long terme de cette approche et que doit-on mettre en œuvre pour parvenir à destination ?

Les professionnels de l'éducation, les administrateurs scolaires et les décideurs devraient envisager d'intégrer les évaluations par le jeu dans leurs systèmes d'évaluation de l'éducation, car elles offrent des avantages uniques par rapport aux approches plus traditionnelles. Les évaluations par le jeu sont spéciales, car elles peuvent refléter les interactions dynamiques, la complexité structurelle et les boucles de rétroaction qui sont présentes dans le monde réel. À long terme, les systèmes d'évaluation intégrés devraient s'appuyer sur des scénarios basés sur le jeu et la simulation pour évaluer la manière dont les élèves intègrent et appliquent leurs connaissances, leurs compétences et leurs aptitudes. Les scénarios peuvent certes concerner des éléments précis d'une matière donnée, mais leur plus grand avantage est peut être de faciliter la mesure des compétences du XXIe siècle comme la résolution de problèmes et la collaboration.

Les avantages des évaluations par le jeu, notamment la capacité d'évaluer des processus cognitifs historiquement difficiles à mesurer, un meilleur alignement avec les programmes de cours modernes et une participation accrue des élèves dans le processus de mesure, en font un élément important de l'avenir de tous les systèmes d'évaluation de l'éducation. Cependant, les approches par le jeu ne produisent souvent pas autant de données chiffrées (des scores) utilisables que nous pourrions l'espérer, surtout si l'on tient compte de leur coût de développement relativement élevé par rapport aux items classiques. Un système d'évaluation rentable et robuste pourrait donc utiliser des scénarios basés sur le jeu en combinaison avec des items d'évaluation classiques et moins coûteux,

qui seraient améliorés grâce à des technologies spécifiques. Une approche pertinente en termes de développement consisterait à utiliser des évaluations classiques relativement peu coûteuses lorsque cela est possible (par exemple, pour mesurer les compétences dans une matière scolaire) et à réserver les évaluations par le jeu plus coûteuses à la mesure de constructions cognitives complexes. De plus, l'utilisation des évaluations par le jeu ne devrait pas se limiter aux seules évaluations sommatives, mais devrait plutôt faire partie d'un système cohérent d'évaluation tout au long de l'année scolaire. Un tel système d'évaluation hybride pourrait en théorie être conçu pour d'autres utilisations, notamment la production de rapports de redevabilité, le pilotage de l'enseignement à l'échelle locale et la modélisation de la progression de chaque élève.

Si l'on veut bénéficier des avantages des évaluations par le jeu et par simulation, les ministères de l'Éducation doivent investir dans les infrastructures qui sont nécessaires à la conception, à la mise en œuvre et à l'administration de ces tests. La mise en œuvre d'une partie de ces capacités peut être sous-traitée à des fournisseurs du secteur privé, mais il faudra également mobiliser des capacités publiques. Il faudra notamment disposer de matériel informatique en quantité suffisante dans les établissements scolaires (bien que l'on ait de plus en plus tendance à envisager des politiques du type « apportez votre propre appareil ») et d'une structure de réseau capable de supporter des vitesses de transfert acceptables.

Références

Arieli-Attali, M., S. Ward, J. Thomas, B. Deonovic et A. von Davier (2019), « The Expanded Evidence-Centered Design (e-ECD) for Learning and Assessment Systems: A Framework for Incorporating Learning Goals and Processes Within Assessment Design », *Frontiers in Psychology*, Vol. 10, http://dx.doi.org/10.3389/fpsyg.2019.00853. [16]

Bergner, Y. et A. von Davier (2018), « Process Data in NAEP: Past, Present, and Future », *Journal of Educational and Behavioral Statistics*, Vol. 44/6, pp. 706-732, http://dx.doi.org/10.3102/1076998618784700. [42]

Bradshaw, L. (2016), « Diagnostic Classification Models », dans *The Handbook of Cognition and Assessment*, John Wiley & Sons, Inc., Hoboken, NJ, USA, http://dx.doi.org/10.1002/9781118956588.ch13. [46]

Braun, H. et A. Kanjee (2006), « Using assessment to improve education in developing nations », dans Braun, H. et al. (dir. pub.), *Improving Education Through Assessment, Innovation, and Evaluation.*, Cambridge, Mass.: American Academy of Arts and Sciences. [1]

Bundesministerium für Bildung und Forschung (s.d.), *ASCOT+*, https://www.ascot-vet.net (consulté le 30 mai 2021). [26]

Chopade, P., D. Edwards, S. Khan, A. Andrade et S. Pu (2019), « CPSX: Using AI-Machine Learning for Mapping Human-Human Interaction and Measurement of CPS Teamwork Skills », *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*, http://dx.doi.org/10.1109/hst47167.2019.9032906. [53]

Chung, G. (2014), « Toward the Relational Management of Educational Measurement Data », *Teachers College Record*, Vol. 116/11. [15]

Ciolacu, M., A. Fallah Tehrani, L. Binder et P. Mugur Svasta (2018), « Education 4.0 - Artificial Intelligence Assisted Higher Education: Early recognition System with Machine Learning to support Students' Success », *2018 IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, http://dx.doi.org/10.1109/siitme.2018.8599203. [49]

Conley, D. (2018), *The Promise and Practice of Next Generation Assessment.*, Harvard University Press, Cambridge, MA. [19]

Cordova, D. et M. Lepper (1996), « Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice », *Journal of Educational Psychology*, Vol. 88/4, pp. 715-730, http://dx.doi.org/10.1037/0022-0663.88.4.715. [34]

Darling-Hammond, L. (2006), « Constructing 21st-Century Teacher Education », *Journal of Teacher Education*, Vol. 57/3, pp. 300-314, http://dx.doi.org/10.1177/0022487105285962. [11]

Darling-Hammond, L., J. Herman, J. Pellegrino, J. Abedi, J. Lawrence Aber, E. Baker, R. Bennett, E. Gordon, E. Haertel, K. Hakuta, A. Ho, R. Lee Linn, P.D. Pearson, J. Popham, L. Resnik, A. Schoenfeld, R. Shalveson, L. Shepard, L. Shulman et C. Steele (2013), « Criteria for High-quality Assessment », *Stanford Center for Opportunity Policy in Education*. [14]

Darrah, M. (2013), « Computer Haptics: A New Way of Increasing Access and Understanding of Math and Science for Students Who are Blind and Visually Impaired », *Journal of Blindness Innovation and Research*, Vol. 3/2, http://dx.doi.org/10.5241/3-47. [44]

de la Torre, J. et J. Douglas (2004), « Higher-order latent trait models for cognitive diagnosis », *Psychometrika*, Vol. 69/3, pp. 333-353, http://dx.doi.org/10.1007/bf02295640. [47]

Deterding, S., D. Dixon, R. Khaled et L. Nacke (2011), « From game design elements to gamefulness », *Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments - MindTrek '11*, http://dx.doi.org/10.1145/2181037.2181040. [38]

DiCerbo, K. (2014), « Game-Based Assessment of Persistence », *Educational Technology & Society*, Vol. 17/1, pp. 17-28. [3]

Duncan, R. et C. Hmelo-Silver (2009), « Learning progressions: Aligning curriculum, instruction, and assessment », *Journal of Research in Science Teaching*, Vol. 46/6, pp. 606-609, http://dx.doi.org/10.1002/tea.20316. [9]

Echeverría, A., M. Améstica, F. Gil, M. Nussbaum, E. Barrios et S. Leclerc (2012), « Exploring different technological platforms for supporting co-located collaborative games in the classroom », *Computers in Human Behavior*, Vol. 28/4, pp. 1170-1177. [48]

Fadel, C., M. Bialik et B. Trilling (2015), *Four-dimensional Education: the Competencies Learners Need to Succeed.*, Center for Curriculum Redesign, Cambridge, MA. [30]

Ferrara, S., E. Lai, A. Reilly et P. Nichols (2017), « Principled Approaches to Assessment Design, Development, and Implementation », dans The *Handbook of Cognition and Assessment,* John Wiley & Sons, Inc., Hoboken, NJ, USA, http://dx.doi.org/10.1002/9781118956588.ch3. [22]

Games for change (s.d.), Games for change, http://www.gamesforchange.org/game/simcityedu-pollution-challenge/ (consulté le 30 avril 2021). [51]

Glasslab (s.d.), *SIMCITY edu: pollution challenge,* https://s3-us-west-1.amazonaws.com/playfully-games/SC/brochures/SIMCITYbrochure_v3small.pdf (consulté le 30 avril 2021). [52]

Gobert, J., R. Baker et M. Wixon (2015), « Operationalizing and Detecting Disengagement Within Online Science Microworlds », *Educational Psychologist,* Vol. 50/1, pp. 43-57, http://dx.doi.org/10.1080/00461520.2014.999919. [37]

Gong, B. (2010), *Using balanced assessment systems to improve student learning and school capacity: An introduction.,* Council of Chief State School Officers, Washington, DC. [2]

Hamari, J., D. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke et T. Edwards (2016), « Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning », *Computers in Human Behavior,* Vol. 54, pp. 170-179, http://dx.doi.org/10.1016/j.chb.2015.07.045. [33]

Klieme, E. (2020), « Policies and Practices of Assessment: A Showcase for the Use (and Misuse) of International Large Scale Assessments in Educational Effectiveness Research », dans *International Perspectives in Educational Effectiveness Research*, Springer International Publishing, Cham, http://dx.doi.org/10.1007/978-3-030-44810-3_7. [25]

Madaus, G. et M. Russell (2010), « Paradoxes of High-Stakes Testing », *Journal of Education,* Vol. 190/1-2, pp. 21-30, http://dx.doi.org/10.1177/002205741019000l-205. [27]

Marion, S., J. Thompson, C. Evans, J. Martineau et N. Dadey (2019), « A Tricky Balance: The Challenges and Opportunities of Balanced Systems of Assessment », dans *Paper Presented at the Annual Meeting of the National Council on Measurement in Education Toronto, Ontario April 6, 2019*, National Center for the Improvement of Educational Assessment, https://www.nciea.org/sites/default/files/inline-files/Marion%20et%20al_A%20Tricky%20Balance_031319.pdf (consulté le 2 janvier 2020). [23]

Martone, A. et S. Sireci (2009), « Evaluating Alignment Between Curriculum, Assessment, and Instruction », Vol. 79/4, pp. 1332-1361, http://dx.doi.org/10.3102/0034654309341375. [17]

Mislevy, R. (2018), *Sociocognitive Foundations of Educational Measurement.,* Routledge, New York. [36]

Mislevy, R., R. Almond, D. Yan et L. Steinberg (2000), « Bayes nets in educational assessment: Where do the numbers come from? », *(CSE Report 518). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).* [45]

Mislevy, R., J. Behrens, K. Dicerbo et R. Levy (2012), « Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining », *Journal of educational data mining,* Vol. 4/1, pp. 11-48. [6]

Mislevy, R., A. Oranje, M. Bauer, A. von Davier, J. Hao, S. Corrigan, E. Hoffman, K. DiCerbo et M. John (2014), *Psychometric Considerations in Game-Based Assessment.,* Glasslab Games, Redwood City, CA. [50]

National Research Council (2001), *What Students Know: The Science and Design of Educational Assessment.,* National Academies Press, Washington, D.C., http://dx.doi.org/10.17226/10019. [13]

Nichols, S. et H. Dawson (2012), « Assessment as a Context for Student Engagement », dans *Handbook of Research on Student Engagement,* Springer US, Boston, MA, http://dx.doi.org/10.1007/978-1-4614-2018-7_22. [12]

Oranje, A., B. Mislevy, M. Bauer et G.Tanner Jackson (2019), « Summative Game-based Assessment », dans Ifenthaler, D. and Y. Kim (dir. pub.), Ga*me-based Assessment Revisited*, Springer. [20]

Pellegrino, J. et M. Hilton (dir. pub.) (2012), *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. Committee on Defining Deeper Learning and 21st Century Skills.*, National Academies Press, Washington, D.C., http://dx.doi.org/10.17226/13398. [18]

Perie, M., S. Marion et B. Gong (2009), « Moving Toward a Comprehensive Assessment System: A Framework for Considering Interim Assessments », *Educational Measurement: Issues and Practice*, Vol. 28/3, pp. 5-13, http://dx.doi.org/10.1111/j.1745-3992.2009.00149.x. [10]

Rose, D. (2000), « Universal Design for Learning », *Journal of Special Education Technology,* Vol. 15/3, pp. 45-49, http://dx.doi.org/10.1177/016264340001500307. [43]

Sabourin, J., J. Rowe, B. Mott et J. Lester (2011), « When Off-Task is On-Task: The Affective Role of Off-Task Behavior in Narrative-Centered Learning Environments », dans *Lecture Notes in Computer Science, Artificial Intelligence in Education*, Springer Berlin Heidelberg, Berlin, Heidelberg, http://dx.doi.org/10.1007/978-3-642-21869-9_93. [40]

Sanders, W. et S. Horn (1995), « Educational Assessment Reassessed », *education policy analysis archives*, Vol. 3, pp. 6, http://dx.doi.org/10.14507/epaa.v3n6.1995. [8]

Seelow, D. (2019), « The Art of Assessment: Using Game Based Assessments to Disrupt, Innovate, Reform and Transform Testing », *Journal of Applied Testing Technology*, Vol. 20/S1, pp. 1-16. [32]

Shaffer, D., D. Hatfield, G. Navoa Svarovsky, P. Nash, A. Nulty, E. Bagley, K. Frank, A. Rupp et R. Mislevy (2009), « Epistemic Network Analysis: A Prototype for 21st-Century Assessment of Learning », *International Journal of Learning and Media*, Vol. 1/2, pp. 33-53, http://dx.doi.org/10.1162/ijlm.2009.0013. [4]

Shepard, L., W. Penuel et J. Pellegrino (2018), « Using Learning and Motivation Theories to Coherently Link Formative Assessment, Grading Practices, and Large-Scale Assessment », *Educational Measurement: Issues and Practice*, Vol. 37/1, pp. 21-34, http://dx.doi.org/10.1111/emip.12189. [21]

Shute, V. (2011), *Stealth assessment in computer-based games to support learning. Computer games and instruction.*, Information Age Publishers, Charlotte, NC, http://myweb.fsu.edu/vshute/pdf/shute%20pres_h.pdf. [5]

Shute, V. et M. Ventura (2013), « Stealth Assessment: Measuring and Supporting Learning in Video Games », dans John, D. and C. MacArthur (dir. pub.), *Foundation Reports on Digital Media and Learning.*, The MIT Press, Cambride, MA, http://dx.doi.org/10.7551/mitpress/9589.001.0001. [35]

Shute, V., M. Ventura, M. Bauer et D. Zapata-Riviera (2009), « Melding the power of serious games and embedded assessment to monitor and foster learning », *Serious games: Mechanisms and effects*, Vol. 2, pp. 295-321. [7]

Snow, E., L. Allen, M. Jacovina et D. McNamara (2015), « Does agency matter?: Exploring the impact of controlled behaviors within a game-based environment », *Computers & Education*, Vol. 82, pp. 378-392, http://dx.doi.org/10.1016/j.compedu.2014.12.011. [41]

Stecher, M. et L. Hamilton (2014), Measuring hard-to-measure student competencies: A research and development plan., RAND Corporation, Santa Monica, CA, https://www.rand.org/pubs/research_reports/RR863.html. [31]

Trilling, B. et C. Fadel (2009), *21st century skills: Learning for Life in Our Times.*, Jossey-Bass. [28]

Verger, A., L. Parcerisa et C. Fontdevila (2019), « The growth and spread of large-scale assessments and test-based accountabilities: a political sociology of global education reforms », *Educational Review*, Vol. 71/1, pp. 5-30, http://dx.doi.org/10.1080/00131911.2019.1522045. [24]

Vincent-Lancrin, S., C. Gonzalez-Sancho, M. Bouckaert, F. de Luca, M. Fernandez-Barrerra, G. Jacotin, J. Urgel et Q. Vidal (2019), *Fostering Students' Creativity and Critical Thinking: What it Means in School*, Educational Research and Innovation, OECD Publishing, Paris, https://dx.doi.org/10.1787/62212c37-en. [29]

Yang, F., L. Leqi, Y. Wu, Z. Lipton, P. Ravilkimar, W. Cohen et T. Mitchel (2019), « Game Design for Eliciting Distinguishable Behavior », *Paper prepared for the 33rd Conference on Neural Information Processing Systems.*, https://papers.nips.cc/paper/8716-game-design-for-eliciting-distinguishable-behavior.pdf (consulté le 2 janvier 2020). [39]

Extrait de :

OECD Digital Education Outlook 2021

Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots

Accéder à cette publication :
https://doi.org/10.1787/589b283f-en

Merci de citer ce chapitre comme suit :

Buckley, Jack*, et al.* (2022), « Des évaluations éducatives par le jeu », dans OCDE, *OECD Digital Education Outlook 2021 : Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, Éditions OCDE, Paris.

DOI: https://doi.org/10.1787/0c4dab5d-fr

OCDE