

Introduction

Ce document vise à discuter des bases de l'évaluation de l'utilité, dans le contexte de la conception du tableau de bord d'un EIAH, et de réalisation de processus d'analyse qui conduisent à la constitution des indicateurs présentés dans ces tableaux de bord (TDB). Précisons que le TDB correspond à ce que l'utilisateur voit classiquement (indicateurs, graphiques), tandis que les processus d'analyse, quelle que soit leur forme, correspondent à l'ensemble des opérations et transformations conduisant à la construction de ces indicateurs. Si ces opérations sont généralement inaccessibles à l'utilisateur final, l'accent est mis dans le consortium Hubble sur leur formalisation et leur réification sous la forme de divers artefacts.

Ainsi, lorsque nous nous intéressons ici à l'utilité, nous devons penser autant à l'utilité des indicateurs et représentations graphiques fournies dans le tableau de bord, qu'à celle de la représentation du processus d'analyse dans son ensemble. Pour que le processus d'analyse soit 'utile', d'autres dimensions que l'utilité perçue par l'utilisateur doivent être prises en considération. Plus particulièrement, le processus d'analyse doit être suffisamment générique pour pouvoir être reproduit sur des jeux de données variés, et les transformations subies par les données doivent être telles qu'elles n'invalident pas la pertinence d'un indicateur ou d'une représentation graphique.

Notons dès à présent que dans le champ de l'éducation les publications sur l'évaluation des tableaux de bord sont relativement rares, et celles sur les processus d'analyse virtuellement inexistantes. On trouve certes quelques travaux relatifs à l'évaluation des TDB hors du champ de l'éducation (Dyczkowski, Korczak, & Dudycz, 2014; Iandoli, Quinto, De Liddo, & Buckingham Shum, 2014; Reinhardt, Mletzko, Drachsler, & Sloep, 2014; Tarrant et al., 2017), et les rares recherches réalisées dans le domaine de l'éducation que nous avons identifiées (Stephens-Martinez, Hearst, & Fox, 2014; Toisoul, 2017; Faber, Luyten, & Visscher, 2017) sont loin de couvrir la diversité des méthodes employées dans le champ de la conception. Aussi, si nous mobiliserons plusieurs exemples de la littérature portant sur l'évaluation de tableaux de bord, notre propos se veut relativement générique.

Dans ce document, nous nous proposons dans une première partie de préciser la définition de l'utilité, de distinguer ses différentes dimensions, et de mettre en perspective cette notion avec celle de besoin. Pour cette partie, nous nous fonderons principalement sur une revue de la littérature consacrée à la question (Loup-Escande, Burkhardt, & Richir, 2013). Nous évoquerons en conclusion le fait que cette approche de l'utilité, centrée sur l'utilité perçue, est insuffisante si l'on s'intéresse aux processus d'analyse tel qu'on peut le faire dans Hubble.

Les deux parties suivantes seront consacrées à une évaluation de l'utilité de tableaux de bord et de processus d'analyse, avec dans un premier temps un focus sur une évaluation théorique de l'utilité, fondée sur la modélisation de l'activité de l'utilisateur. L'évaluation de l'utilité, si elle doit intervenir le plus tôt possible dans le processus de conception, se doit d'être

précédée d'une analyse des besoins, et d'une modélisation de la situation dans laquelle s'insère le tableau de bord. Cette démarche relève de la dimension analytique de l'évaluation.

Dans une dernière partie, nous nous penchons sur quelques techniques relevant de l'évaluation empirique des tableaux de bords et des processus d'analyse, en prenant notamment quelques exemples issus de la littérature sur les tableaux de bord des MOOC. Avant d'entrer dans le vif du sujet, notons que l'on considérera ici que le lecteur est familier avec les notions d'utilisabilité et d'acceptabilité (Tricot et al., 2003), ainsi qu'avec les notions d'évaluation analytique et d'évaluation empirique, de test exploratoire ou de test d'évaluation (Bastien & Tricot, 2015). Elles ne seront pas redéfinies ou traitées dans ce document, pour nous permettre de nous concentrer davantage sur la notion d'utilité.

I. Evaluation de l'utilité et analyse des besoins

I.1 Définition de l'utilité et de ses différentes dimensions

Commençons par nous attacher aux questions de définition. Alors que l'utilisabilité caractérise l'interface, "l'utilité est une propriété se situant à l'interface entre le système, le sujet et la situation, correspondant à une réponse à un besoin ou à une attente de l'utilisateur. [...] D'autres termes aux significations proches ou partiellement connexes à l'utilité sont parfois aussi mobilisés, par exemple celui de pertinence, ou encore d'adaptabilité et de compatibilité (Scapin & Bastien, 1997)". Nous conserverons dans ce document le terme d'utilité.

Loup-Escande et al. (2013) soulignent que beaucoup des réflexions menées sur l'ergonomie portent sur l'utilisabilité, "à tel point qu'il n'est pas rare que l'ergonomie soit assimilée à cette seule dimension de l'utilisabilité (Burkhardt & Sperandio, 2014). Pour certains auteurs, toutefois, le critère d'utilité est considéré comme premier. L'idée est en effet que même parfaitement utilisable, un objet inutile peut difficilement être qualifié d'ergonomique (Nielsen, 1993; Burkhardt & Sperandio, 2004)".

La question de l'évaluation de l'utilité est compliquée par le caractère davantage relatif des critères utilisés pour réaliser l'évaluation, contrairement. Comme le souligne toujours Loup-Escande et al. (2013), toute tentative se heurte au fait « qu'il n'y a pas aujourd'hui – pour le praticien comme pour le chercheur en ergonomie – de cadre théorique ou méthodologique traitant explicitement de l'utilité pour guider la conception et définir les objectifs à atteindre, comparativement à la précision et la plus grande diffusion des cadres, des méthodes et des objectifs associés à l'utilisabilité. ».

Dans leur réflexion sur la question, les auteurs distinguent par la suite deux dimensions de l'utilité : l'*utilité-destination* et l'*utilité-valeur*. "La première acception correspond aux caractéristiques fonctionnelles ou non-fonctionnelles de l'artefact, alors que la seconde correspond aux améliorations ou aux bénéfices significatifs que peut apporter l'artefact aux

utilisateurs. [...] l'utilité-destination renvoie explicitement à la description des fonctions du système (Nielsen, 1993) et des utilisations inscrites dans l'artefact."

Nielsen (1993) distingue quant à lui, d'un côté, l'utilité désignant les buts que le système est censé permettre d'atteindre (« utility », traduite par Tricot et al. (2003) au travers des termes « utilité théorique »), et, de l'autre côté, l'utilité désignant le but que le système, utilisé par un utilisateur donné, permet effectivement d'atteindre dans des situations spécifiques (« usefulness », c'est-à-dire l'utilité réelle / pratique ; Tricot et al. 2003)."

Les deux dimensions d'utilité théorique et d'utilité pratique ne sont pas indépendantes, au sens où l'utilité pratique inclut l'utilité théorique en plus de l'utilisabilité. Les méthodologies associées à l'évaluation de ces dimensions néanmoins distinctes ne sont pas tout à fait distinctes, l'utilité pratique invitant davantage à l'emploi des méthodes de l'évaluation empirique. Ces précisions étant données, il nous faut maintenant mettre en perspective la notion d'utilité et la notion de besoin. Les deux termes correspondent à une emphase sur des pôles distincts du triplet artefact/ utilisateur/situation, l'utilité étant associée davantage à l'artefact et le besoin davantage à l'utilisateur.

I.2 Besoins conscients, non conscients et latents

Trois dimensions décrivant les besoins nous semblent utiles à rapporter ici. La première dimension est le degré d'explicitation des besoins, le second est son caractère fonctionnel ou non, le dernier correspond à son degré de généralité. Nous nous focaliserons ici uniquement sur la première. À cette fin, il faut tout d'abord distinguer les besoins conscients, des besoins non-conscients et des besoins latents. Les besoins conscients sont explicitement formulés par les utilisateurs (Robertson, 2001), spontanément ou non. Les entretiens individuels comme les focus groupes peuvent être des méthodes privilégiées pour la récolte de tels besoins. Néanmoins, "les besoins conscients sont généralement incomplets, voire en décalage avec les besoins réels," (Robertson, 2001), d'où les limites des méthodes fondées uniquement sur l'approche autodéclarative.

Les besoins non conscients "résultent de plusieurs facteurs possibles (Robertson, 2001). L'absence d'opportunité offerte aux utilisateurs concernés pour réfléchir à leur activité actuelle est un premier facteur. Un second réside dans la méconnaissance qu'ont les utilisateurs des propriétés et des potentialités offertes par la technologie choisie, empêchant ainsi la mise en correspondance de ces potentialités avec les propriétés critiques de la tâche et de l'activité ciblée. Un autre facteur majeur réside dans la nature non verbalisable de certains besoins, liée notamment à une grande part d'automatismes dans l'activité". Les techniques reposant sur l'observation des utilisateurs en conditions écologiques peuvent être mobilisées de manière privilégiée pour identifier ces besoins non conscients. Elles s'inscrivent dans une vision de la conception que l'on qualifie parfois "d'empathic design", ou design empathique.

Enfin, les besoins « latents » se « caractérisent par leur nature non encore avérée ou encore « inimaginée » (Robertson, 2001). Ce troisième type de besoin est un enjeu important pour les technologies émergentes, encore au stade du développement en laboratoire, et à la recherche de leurs applications.” Nous sommes précisément dans cette situation lorsque l’on conçoit un tableau de bord pour un EIAH, ou que l’on cherche à réifier un processus d’analyse dans un artefact. Il nous faut maintenant préciser le lien qui unit l’analyse des besoins, l’utilité, et l’évaluation de cette dernière. “Le besoin peut désigner initialement les fonctions ou propriétés possibles, voire souhaitables pour l’artefact à concevoir.”. Ce n’est que lorsqu’il a été traduit sous la forme de fonctions et de propriétés de l’artefact que l’utilité de ces dernières peut-être évaluée.

L’utilité réelle d’un système découle donc de la traduction pertinente des besoins des utilisateurs au niveau de l’artefact, sans toutefois s’y réduire pour au moins deux raisons. D’une part, les besoins sont susceptibles d’évoluer du fait de l’évolution du contexte et de l’apprentissage des utilisateurs. D’autre part, l’artefact et son environnement peuvent à leur tour révéler ou susciter des besoins inédits parmi les utilisateurs. Ces diverses considérations étant établies, penchons-nous maintenant sur la question des méthodes d’évaluation de l’utilité, en prenant dans un premier temps le point de vue de l’évaluation théorique.

II Évaluation analytique de l’utilité

Dans une approche de type « évaluation analytique », le diagnostic ergonomique se fonde sur une modélisation de l’activité instrumentée, que l’artefact soit le tableau de bord ou toute réification du processus d’analyse. L’idée est ici de faire une analyse de l’adéquation entre les besoins identifiés, le scénario d’utilisation de l’artefact, et le contenu du tableau de bord ou du processus d’analyse. Ce travail implique dans un premier temps de modéliser l’activité de l’utilisateur, afin de mieux appréhender comment va s’y insérer l’artefact (Haradji & Faveaux, 2006). La légitimité de cette approche repose sur le fait qu’il est d’autant plus coûteux d’identifier un dysfonctionnement que cette identification a lieu tardivement dans le processus de conception, à l’occasion d’une évaluation empirique.

II.1 Modélisation de l’activité non instrumentée par l’application

Comme le soulignent Haradji & Faveaux (2006), dont l’article va guider notre réflexion, « la connaissance de l’activité passe par une analyse en situation réelle de travail. Elle ne peut pas être abordée uniquement hors situation (les interviews) ou à partir d’une situation expérimentale qui réduirait le travail à quelques variables isolées ». La modélisation de l’activité passe dans un premier temps par une observation de son activité, en situation écologique, observation qui peut être complétée par des entretiens permettant de mieux appréhender comment il raisonne. En effet, pour les auteurs « le point de vue de l’acteur semble un point de vue préférable à tout autre quand il s’agit d’interpréter son activité (Nogry, Jean-Daubias, & Ollagnier-Beldame, 2004). »

Les auteurs fondent leur description de l'activité sur deux notions complémentaires, la notion d'unité significative pour l'acteur. Il s'agit ici de 'mettre en évidence des cohérences d'actions, de communications, d'interprétations, etc., qui sont significatives pour l'utilisateur.'

De cette notion d'unité significative, découle des unités élémentaires, qui correspondent à un degré de définition plus fin de l'activité, nommées unités élémentaires d'interaction, qui peuvent être agrégées en séquences d'interaction, puis en macro-séquences d'interaction. Cette notion d'unité significative est complétée par celle de récit réduit partiel, qui "correspond donc à un enchaînement d'unités significatives plus ou moins larges présentant ainsi, sous forme de récit, un moment d'activité de l'utilisateur". Nous nous proposons d'illustrer dans le tableau ci-dessous une macro-séquence d'interaction correspondant au récit partiel de l'activité d'un enseignant en l'absence d'une instrumentation spécifique (Tableau 1).

Macro-séquence d'interaction	Séquence d'interaction	Unité élémentaire d'interaction
Faire participer l'ensemble de la classe à une séquence de mutualisation d'idées	Demander aux élèves de rédiger des idées	Donner une consigne aux élèves pour qu'ils génèrent des idées
		Passer entre les rangs pour vérifier que la consigne est bien suivie
	Catégoriser des idées	Mutualiser au tableau les idées produites par les élèves
		Demander à la classe de produire des catégories pertinentes pour ranger ces idées
		Faire catégoriser par la classe les idées ainsi mutualisées

Tableau 1. Récit partiel réduit lié à un moment d'activité de l'enseignant

Le modèle d'activité est alors construit à partir d'une généralisation de récits partiels réduits menés auprès de plusieurs types d'utilisateurs. On passe ensuite du modèle de l'activité à un modèle de la tâche de la future application.

II.2 Du modèle de l'activité au modèle de la tâche future

La logique de la construction du modèle de la tâche future est sensiblement différente, puisque l'on passe d'une logique de description de l'activité existante à la logique de définition d'une activité qui n'existe pas encore. Il s'agit dès lors d'identifier les "logiques à préserver, et les dysfonctionnements à dépasser" (Haradji & Faveaux, 2006). Cette approche impose d'appréhender les mutations de l'activité. Toujours pour les auteurs, "le passage de la description des tâches de l'existant à la description des tâches futures permet de poser explicitement les transformations qui sont envisagées dans le métier (par exemple, du fait des évolutions techniques) et d'en imaginer les conséquences sur l'activité future". Nous proposons de reproduire dans le tableau suivant un modèle de tâches futures qui intégrerait l'utilisation par l'enseignant d'un tableau de bord lors d'une séance de génération et de catégorisation d'idées.

Macro-séquence d'interaction	Séquence d'interaction	Unité élémentaire d'interaction
Faire participer l'ensemble de la classe à une séquence de mutualisation d'idées	Demander aux élèves de rédiger des idées	Donner une consigne aux élèves pour qu'ils génèrent des idées
		Identifier via le tableau de bord les élèves ou les groupes n'ayant pas participé à la génération d'idées
		Encourager via l'application les élèves ou les groupes n'ayant pas participé à produire des idées

Tableau 2. Modèle des tâches futures de l'utilisateur, pour le tableau de bord

Ancrer l'évaluation de l'utilité dans la phase de spécification nous semble indispensable si l'on souhaite que l'artefact permette la réalisation de cette tâche future. Nous ne nous attarderons pas ici sur le type de modèle que l'on peut mobiliser pour décrire l'activité de l'utilisateur ni celle de l'application, et laisserons au lecteur le choix d'identifier les modèles pertinents eu égard à ses problématiques.

En ce qui concerne les processus d'analyse tels que conçus dans Hubble, diverses approches peuvent être suivies si l'on s'intéresse à l'utilité théorique. Tout d'abord, on peut adopter une définition de l'utilité équivalente à celle que nous avons adoptée jusqu'ici, et suivre par conséquent une approche similaire à celle que nous avons décrite pour le tableau de bord. Le détail d'un processus d'analyse est plus susceptible d'intéresser un entrepreneur, ou un chercheur, qu'un enseignant ou un élève, mais le principe reste le même. Une telle évaluation analytique de l'utilité implique une réification du processus d'analyse dans un artefact – synopsis, carte mentale interactive, processus d'analyse narré, etc. – car l'on ne peut évaluer ici l'utilité que d'un artefact. On s'attachera ensuite à représenter un modèle des tâches futures dans lequel s'insère ce processus d'analyse réifié. On ne peut néanmoins, dans le cas particulier des processus d'analyse, se cantonner à cette vision de l'utilité issue du diagnostic ergonomique, car d'autres dimensions doivent être prises en considération, que nous allons aborder maintenant.

II.3 Une évaluation de l'utilité des processus d'analyse

Comme nous le soulignons en introduction de ce document, on ne peut se cantonner à l'acception de l'utilité telle que définie dans la littérature sur le diagnostic ergonomique, et nous proposons de nous pencher sur au moins deux dimensions du processus d'analyse pour compléter l'évaluation de son utilité. La première est son niveau de généralité. Le processus doit être suffisamment instancié pour être opératoire, mais suffisamment générique pour pouvoir être transféré facilement à une classe de situations hétérogènes. La seconde est relative aux transformations subies par les données.

Généricité du processus d'analyse

« Ce paragraphe ne relève pas des sciences humaines, mais plutôt de l'informatique. Je n'ai pas le background requis pour le rédiger correctement »

Transformations subies par les données

L'utilité d'un indicateur relatif à l'activité d'utilisateur peut être mise à mal par certaines transformations subies par les données au cours du processus d'analyse. L'analyse du processus d'analyse doit être en mesure de révéler si de tels problèmes peuvent survenir du fait de ces transformations. Selon ce point de vue, le type d'artefact dans lequel s'incarne le processus d'analyse importe peu, contrairement à l'approche par modélisation de l'activité que nous avons proposée plus tôt. Il convient avant tout de comprendre, pour chaque étape du processus d'analyse, quelles sont celles qui sont susceptibles d'entamer l'utilité de l'indicateur ou de la représentation graphique.

« Il faudrait sans doute ici une approche plus générique des transformations subies par les données. »

III Évaluation empirique de l'utilité

La dernière partie de ce document est consacrée à quelques techniques d'évaluation empirique de l'utilité. Nous évoquerons deux familles de techniques, les méthodes qualitatives, qui incluent notamment diverses techniques d'observation ainsi que la réalisation d'entretiens individuels ou collectifs, et les questionnaires. Nous discuterons en conclusion de cette partie de quelques approches généralement utilisées dans les instruments dédiés à l'évaluation empirique de tableaux de bord. À nouveau, nous distinguerons l'évaluation des tableaux de bord, par des enseignants, des élèves, ou des praticiens, et l'évaluation du processus qui a conduit à la création des indicateurs et graphiques du tableau de bord. Par exemple, des utilisateurs comme des chercheurs ou des entreprises peuvent être amenés à ouvrir la boîte noire d'une analyse. Il ne s'agit pas alors d'identifier si ces utilisateurs jugent pertinent tel ou tel indicateur, mais également d'évaluer le processus d'analyse qui a conduit à leur construction.

III.1 Observations, entretiens, questionnaires

Les méthodes qualitatives viennent combler certains des défauts des méthodes quantitatives, et offrent des outils pour appréhender la globalité du phénomène étudié. Les remarques qui suivent, relatives à ces méthodes, sont issues avant tout des articles de Nogry, Jean-Daubias, & Ollagnier-Beldame (2004) et de Jamet (2006). De multiples méthodes qualitatives peuvent être suivies pour réaliser une évaluation de l'utilité de l'artefact qu'est le TDB, ou le processus d'analyse. L'approche - observation, entretiens, questionnaires, etc. - est commune aux deux types d'objet. En revanche, une différenciation des instruments s'impose dès lors que l'on s'intéresse davantage à des composantes du TDB ou au processus d'analyse lui-même.

III.1.1 Observation, entre méthodes ethnographiques et analyse de parcours

Observation et méthodes ethnographiques

L'observation peut être **expérientielle** ou **systématique** si une grille d'observation est mise en place pour coder, le plus souvent en temps réel, les événements qui sont recensés (Postic & Ketele, 1988). On distingue les grilles selon qu'elles sont basées sur un **système de catégories** ou un **système de signes**. Dans le premier cas, l'ensemble des événements observés doivent être rangés dans une catégorie. Dans le second cas, on se focalise uniquement sur des événements signifiants, tels que des incidents éventuels, qui sont eux-mêmes classifiés. Si nous ne sommes pas parvenus à trouver d'exemples d'instruments mobilisés pour l'évaluation de TDB, on pourra se reporter à l'exemple fourni par Strebelle & Depover (2013) pour l'évaluation d'un EIAH instrumentant l'apprentissage collaboratif de la modélisation scientifique. Les auteurs codent tant les événements se déroulant sur l'application que les interactions se déroulant hors de l'application. La démarche pourra être utilisée pour l'évaluation d'un tableau de bord.

Le **recueil des verbalisations** consiste à « demander à l'utilisateur de penser à haute voix durant la tâche proposée afin d'identifier les raisonnements qu'il met en œuvre pour réaliser la tâche demandée. » En d'autres termes, il s'agit d'identifier les raisons qui poussent un élève ou un enseignant à utiliser telle ou telle fonctionnalité. Ce recueil peut avoir lieu au cours de l'observation, ou en aval de celle-ci. Comme le souligne Jamet (2006) dans sa réflexion sur les **protocoles verbaux**, « Ces verbalisations peuvent être concurrentes à la tâche mais elles ralentissent le traitement. Elles peuvent être différées après la tâche, par exemple en diffusant la vidéo du protocole au sujet et en lui demandant de la commenter. ».

Analyse de parcours, et utilisation des traces d'interaction

Notons que les techniques d'observation où l'enquêteur est présent lors de l'utilisation présentent un certain nombre de limites, dont notamment les contraintes de temps, qui empêchent de réaliser un nombre élevé d'observations, ou l'influence de la présence de l'observateur sur l'utilisation de l'artefact. On peut donc envisager des techniques d'observation fondées sur l'analyse de parcours, qui permettent d'envisager des tests utilisateurs de plus grande ampleur, et/ou n'imposant pas la présence de l'enquêteur. Si la fonction peut être identique, on sort ici néanmoins du cadre des méthodes qualitatives.

Les techniques d'analyse des parcours visent à documenter et analyser la trajectoire de l'utilisateur au sein de l'interface qu'est le tableau de bord, ou, plus largement, au sein de l'application qui comprend le tableau de bord. Dans ce dernier cas, l'objectif peut consister à mieux comprendre comment le tableau de bord s'insère dans un usage plus général de l'application. L'analyse des parcours peut se fonder sur des enregistrements vidéos de l'usage de l'interface, réalisés ou non par des **logiciels de capture d'écran** (on parle alors parfois de *screenflow*), ou sur des **analyses de traces d'interaction**, si celles-ci sont disponibles. Cette dernière approche peut être intéressante si les traces reflètent de manière précise le comportement de l'utilisateur, car elle économise un temps conséquent si l'alternative consiste à coder manuellement des événements capturés sur une vidéo. Par ailleurs, la collecte des vidéos de nombreux terminaux, dans de multiples classes, peut être complexe sur le plan logistique.

On pourra citer pour illustrer cette approche le cas d'une expérience menée par Faber et al. (2017), où l'analyse de traces d'interaction a permis d'évaluer l'utilisation d'un tableau de bord dans un contexte où plusieurs centaines d'utilisateurs testaient l'EIAH, dans près de 80 établissements. Un tel travail aurait été presque impossible à mener s'il avait fallu analyser le flux vidéo de dizaines d'utilisateurs. Le succès de l'expérience tient en grande partie en fait qu'il ne s'agissait pas d'appréhender uniquement la navigation au sein du tableau de bord, mais l'utilisation du tableau de bord, entre autres composantes de l'EIAH, testé en conditions écologiques, i.e. sans observateur. Il faut néanmoins prendre garde au fait que les événements les plus signifiants d'une navigation au sein d'un tableau de bord peuvent ne pas générer de traces. À cet égard, une analyse qualitative de l'activité permettant d'identifier les tâches dont les traces sont les plus importantes, peut être requis pour s'assurer que les traces importantes sont collectées.

III.1.2 Entretiens et questionnaires

Les entretiens peuvent faire suite à des observations, ou tout du moins à des utilisations en situation écologique du tableau de bord. On distingue les entretiens selon qu'ils sont individuels ou collectifs, qu'ils sont ouverts, structurés ou semi-structurés, selon le degré de directivité de l'enquêteur dans la conduite de l'entretien. Dans le cas de l'entretien structuré, les questions sont posées dans le même ordre, et formulées de la même manière, pour l'ensemble des participants, contrairement aux autres cas. Comme le soulignent Nogry et al. (2004) « Dans le cadre d'une évaluation en classe, il peut être utile de faire en complément des entretiens avec les enseignants pour confronter les différents points de vue sur le déroulement de l'évaluation. ». Si l'instrument servant de base à l'entretien structuré peut offrir de nombreuses convergences avec un questionnaire, il présente l'intérêt de collecter des réponses ouvertes bien plus riches et développées que ne le permettrait de le faire le même questionnaire.

L'entretien peut être complété par des questionnaires, dont l'objet est de mesure est de mesurer, via des échelles normées, l'utilité perçue de différentes composantes du tableau de bord. Le questionnaire peut être administré en présence de l'enquêteur comme à distance. Dans ce dernier cas, il nous faut évoquer le cas où des maquettes d'interface sont présentées à des utilisateurs potentiels, en particulier dans le cas des tests exploratoires. Ces utilisateurs doivent alors se prononcer sur ces interfaces via des mesures subjectives. Plutôt que de nous attarder sur les variables permettant de caractériser un entretien ou un questionnaire, focalisons-nous maintenant sur le détail des questions pouvant être mobilisées pour réaliser l'évaluation empirique d'un tableau de bord.

III.2 Instruments mobilisés pour l'évaluation du tableau de bord

Les instruments utilisés au cours d'un entretien ou dans une enquête peuvent être relativement similaires, aussi nous ne les distinguerons pas outre mesure dans les paragraphes qui suivent. Notre propos concerne ici uniquement le tableau de bord, les processus d'analyse faisant l'objet d'une réflexion spécifique.

III.2.1 Évaluation du tableau de bord

Les questionnaires visent à évaluer perçue l'utilité du tableau de bord selon plusieurs dimensions définies par l'évaluateur, celles-ci pouvant être spécifiques de la tâche que le TDB doit instrumenter. Nous verrons également que certains auteurs cherchent à développer des instruments génériques visant à évaluer des TDB relativement indépendamment de la tâche à instrumenter. Nous allons sélectionner quelques exemples issus de la littérature sur les MOOC. Notre propos n'est pas ici générique, et nous ne fournissons pas de règles précises pour les construire. Nous nous cantonnons à énumérer quelques approches issues de la littérature susceptibles d'inspirer la construction d'un instrument.

Le premier exemple que nous avons choisi est fondé sur un test exploratoire (Stephens-Martinez et al., 2014), où il s'agit pour l'enquêté d'estimer, via des maquettes, l'utilité de différentes composantes d'un tableau de bord. Le questionnaire inclut des informations sur l'utilisateur (expérience dans l'enseignement, âge, etc.), afin d'interpréter d'éventuelles variations dans les réponses fournies. Les enquêteurs présentent ensuite à l'utilisateur une maquette de l'interface (Figure 1), qu'ils demandent d'évaluer selon plusieurs principes.

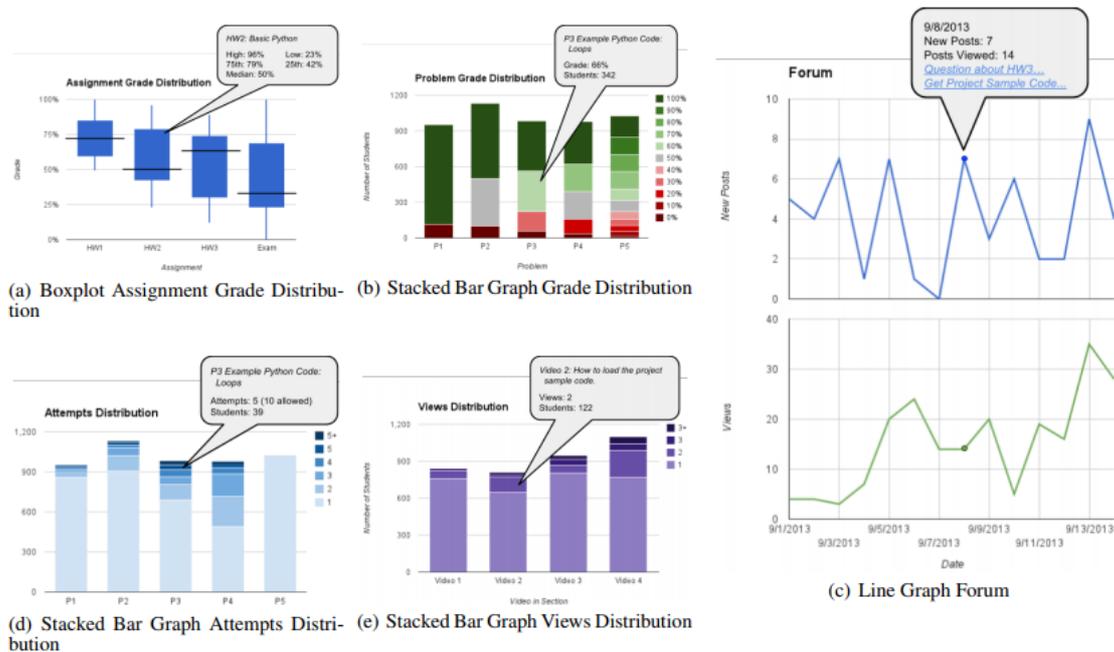


Figure 2. Mockups shown to survey participants. The call out bubbles represents what will be seen if the user hovers the mouse over a part of the graph.

figure 1. Une maquette de tableau de bord présentée aux enquêtés

Les auteurs ont utilisé pour évaluer l'intention d'usage des différentes composantes de ce tableau de bord une échelle (Figure 2), où les enquêtés doivent estimer s'ils l'utilisent déjà la sous-composante du tableau de bord, s'ils utiliseraient si d'aventure celle-ci était disponible, ou s'ils ne l'utiliseraient pas.

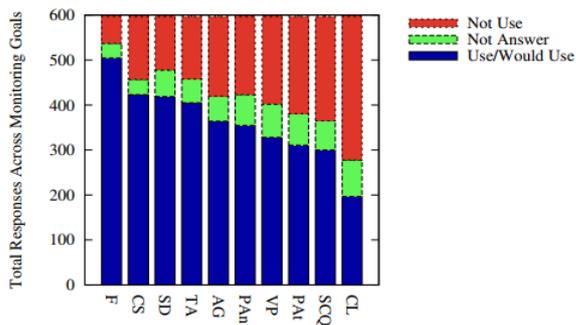


Figure 4. For each information source, the number of participants who use it or would use if available (combined into a single category), do not use, or did not answer. Did not answer means that the participant chose an answer for a subset of the information sources for that Monitoring Goal. It is shown here to see the relative rate of responses.

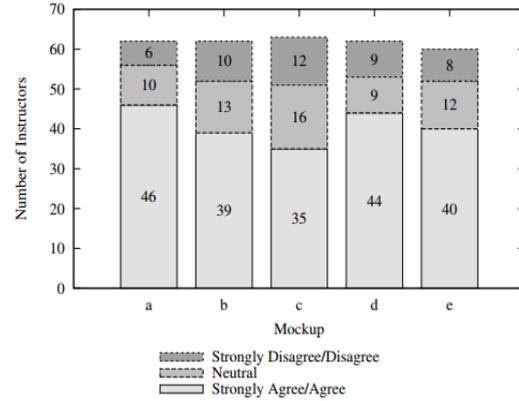


Figure 6. Likert scale responses to the statement "This visualization is useful." Letters correspond to Figure 2's subfigures.

Figure 2. Analyse des réponses sur l'intention d'usage de différentes composantes d'un tableau de bord. On notera l'existence d'approches binaires, où l'utilisateur potentiel se voit demander s'il utiliserait ou non la composante (Gauche). Alternativement, une échelle de valuation pourra être utilisée pour mesurer l'utilité perçue de telle ou telle composante. (Droite)

Certains auteurs préfèrent, plutôt que d'interroger directement les utilisateurs sur la question de l'utilité d'une composante, définir des dimensions génériques pour tous tableaux de bord, comme la clarté des données collectées, de l'intention sous-jacente à la collecte, l'impact sur la motivation et la réflexion (Scheffel, Drachsler, Toisoul, Ternier, & Specht, 2017). Nous donnons un exemple de l'instrument nommé EFLA ci-dessous (Tableau 3), et de son utilisation pour l'évaluation comparative de deux versions successives d'un widget mobilisé dans le cadre d'une plate-forme de MOOC. (Figure 3)

Table 1. Dimensions and items of the learner and the teacher section of the EFLA.

EFLA items for learners/teachers	
Data:	D1 For this LA tool it is clear what data is being collected D2 For this LA tool it is clear why the data is being collected D3 For this LA tool it is clear who has access to the data
Awareness:	A1 This LA tool makes me aware of my/my students' current learning situation A2 This LA tool makes me forecast my/my students' possible future learning situation given my/their (un)changed behaviour
Reflection:	R1 This LA tool stimulates me to reflect on my past learning/teaching behaviour R2 This LA tool stimulates me to adapt my learning/teaching behaviour if necessary
Impact:	I1 This LA tool increases my motivation to study/teach I2 This LA tool stimulates me to study/teach more efficiently I3 This LA tool stimulates me to study/teach more effectively

Tableau 3. Un instrument mobilisé pour l'évaluation comparative de l'utilité de deux versions successives d'un tableau de bord dans un MOOC

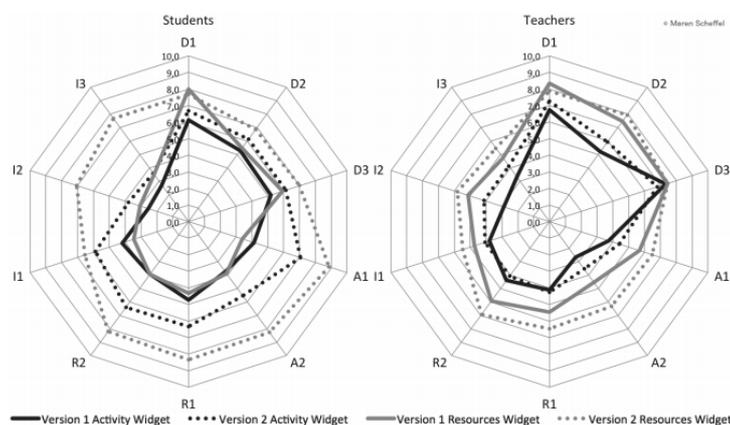


Fig. 1. Average scores of the EFLA items for students (left) and teachers (right) for both versions of both widgets.

Figure 3. Visualisation des résultats de l'évaluation pour deux versions du widget évalué par Scheffel et al. (2017)

Néanmoins, notons que les dimensions choisies dans cet instrument sont relativement peu orthogonales, et qu'un travail d'adaptation est sans doute nécessaire à une utilisation préalable. La logique est relativement similaire dans cette autre étude menée sur le tableau de bord LOCO-Analyst (Ali, Hatala, Gašević, & Jovanović, 2012), les auteurs présentent un ensemble de questions offrant aux participants des réponses ouvertes (Tableau 4), qui sont ensuite codées à la main par les enquêteurs.

Category	Question Description in the questionnaire
Perceived usefulness of the tool for improving the course content/instruction	Q1 [†] : The tool enables me to get an insight into the students' interactions with the learning content
	Q2 [†] : The information the tool provides helps me identify what needs to be improved in the learning content
	Q3 [†] : The tool provides relevant information regarding the students' interactions within the online learning environment.
	Q4 [†] : The information provided by the tool helps me determine how to improve the students' online interactions.
	Q5: The tool helps me identify the domain topics the students were having difficulties with.

Tableau 4 Un questionnaire visant à évaluer deux versions successives du tableau de bord LOCO-Analyst

Pour conclure, soulignons l'existence dans la littérature scientifique d'une autre approche de l'évaluation de l'utilité des composantes d'un tableau de bord, qui nécessite une analyse préalable de l'activité de l'utilisateur. Dans l'exemple qui suit, l'utilisateur doit évaluer l'utilité d'une composante en fonction des tâches qu'il effectue. Dans le cas particulier représenté ci-dessous, l'utilisateur doit déterminer à quelle étape de la préparation de leur cours ils pensaient qu'une telle visualisation leur serait la plus utile (Figure 4).

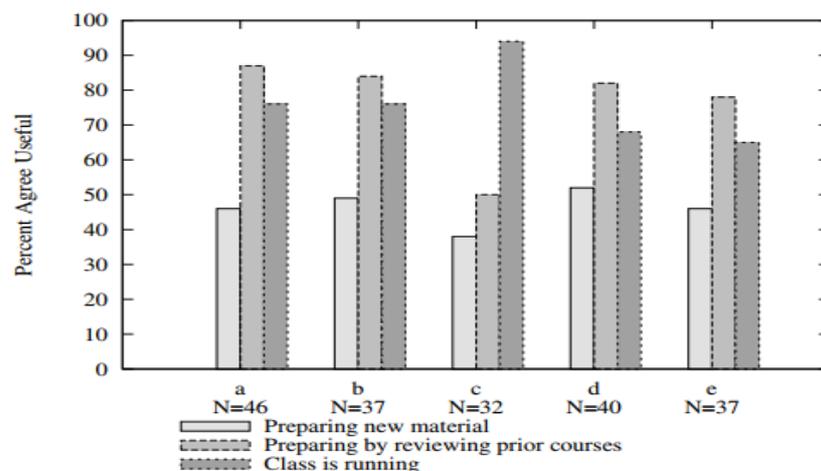


Figure 8. Responses to the question “When would you use this mockup,” with choices in terms of three phases in a “MOOC cycle”: preparing new material, preparing by reviewing previous course runs, and while the class is running. Letters correspond to Figure 2's subfigures.

Figure 4. L'utilité de différentes composantes du tableau de bord, perçue au travers de la phase de l'activité au cours de laquelle l'utilisateur envisage de mobiliser l'artefact

De nombreuses autres publications de cet acabit sont disponibles dans la littérature scientifique, mais leur recensement dépasse les objectifs de ce document. Nous proposons pour conclure de nous attarder sur des instruments mobilisables pour l'évaluation du processus d'analyse.

III.2.2 Instruments mobilisés par l'évaluation du processus d'analyse

« Je pense que la logique est très différente de celle de l'évaluation des tableaux de bord, et je me demande même si cela relève du champ des sciences humaines. J'avoue que je sèche »

Références

- Ali, L., Hatala, M., Gašević, D., & Jovanović, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education*, 58(1), 470–489. <https://doi.org/10.1016/j.compedu.2011.08.030>
- Bastien, J. M. C., & Tricot, A. (2015). *L'évaluation ergonomique des documents électroniques*. Presses Universitaires de France. Retrieved from <http://www.cairn.info/revue-le-travail-humain-puf-2008--p-205.htm>
- Burkhardt, J.-M., & Sperandio, J.-C. (2014). 26. *Ergonomie et conception informatique*. Presses Universitaires de France. Retrieved from <https://www.cairn.info/ergonomie--9782130514046-page-437.htm>
- Dyczkowski, M., Korczak, J., & Dudycz, H. (2014). Multi-criteria evaluation of the intelligent dashboard for SME managers based on scorecard framework. In *2014 Federated Conference on Computer Science and Information Systems* (pp. 1147–1155). <https://doi.org/10.15439/2014F388>
- Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education*, 106, 83–96. <https://doi.org/10.1016/j.compedu.2016.12.001>
- Haradji, Y., & Faveaux, L. (2006). Évolution de notre pratique de conception (1985-2005). *Activités*, 3(1). <https://doi.org/10.4000/activites.1852>
- Iandoli, L., Quinto, I., De Liddo, A., & Buckingham Shum, S. (2014). Socially augmented argumentation tools: Rationale, design and evaluation of a debate dashboard. *International Journal of Human-Computer Studies*, 72(3), 298–319. <https://doi.org/10.1016/j.ijhcs.2013.08.006>
- Jamet, E. (2006). Une présentation des principales méthodes d'évaluation des EIAH en psychologie cognitive. *Sciences et Technologies de l'Information et de La Communication Pour l'Éducation et La Formation (STICEF)*, 13, 13 pages.
- Loup-Escande, É., Burkhardt, J.-M., & Richir, S. (2013). Anticiper et évaluer l'utilité dans la conception ergonomique des technologies émergentes : une revue, Anticipating and evaluating the usefulness of emerging technologies in ergonomic design: a review of usefulness in design. *Le travail humain*, 76(1), 27–55.
- Nielsen, J. (1993). *Usability Engineering*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Nogry, S., Jean-Daubias, S., & Ollagnier-Beldame, M. (2004). Évaluation des EIAH : une nécessaire diversité des méthodes (pp. 265–271). Presented at the Technologies de l'Information et de la Connaissance dans l'Enseignement Supérieur et l'Industrie, Université de Technologie de Compiègne. Retrieved from <https://edutice.archives-ouvertes.fr/edutice-00000729/document>
- Postic, M., & Ketele, J.-M. D. (1988). *Observer les situations éducatives*. Paris: Presses Universitaires de France - PUF.
- Reinhardt, W., Mletzko, C., Drachsler, H., & Sloep, P. B. (2014). Design and evaluation of a widget-based dashboard for awareness support in Research Networks. *Interactive*

- Learning Environments*, 22(5), 594–613.
<https://doi.org/10.1080/10494820.2012.707126>
- Robertson, S. (2001). Requirements trawling: techniques for discovering requirements. *International Journal of Human-Computer Studies*, 55(4), 405–421.
<https://doi.org/10.1006/ijhc.2001.0481>
- Scapin, D. L., & Bastien, J. M. C. (1997). Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology*, 16(4–5), 220–231. <https://doi.org/10.1080/014492997119806>
- Scheffel, M., Drachsler, H., Toisoul, C., Ternier, S., & Specht, M. (2017). The Proof of the Pudding: Examining Validity and Reliability of the Evaluation Framework for Learning Analytics. In *Data Driven Approaches in Digital Education* (pp. 194–208). Springer, Cham. https://doi.org/10.1007/978-3-319-66610-5_15S
- Stephens-Martinez, K., Hearst, M. A., & Fox, A. (2014). Monitoring MOOCs: Which Information Sources Do Instructors Value? In *Proceedings of the First ACM Conference on Learning @ Scale Conference* (pp. 79–88). New York, NY, USA: ACM. <https://doi.org/10.1145/2556325.2566246>
- Strebelle, A., & Depover, C. (2013). Analyse d'activités collaboratives à distance dans le cadre d'un dispositif d'apprentissage de la modélisation scientifique. *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, 1(3). <https://doi.org/10.4000/dms.324>
- Tarrant, C., Carr, S., Kirtley, J., Mukherjee, T., Shaw, L., & Willars, J. (2017). *Evaluation of an Education Quality Dashboard (EQD) in a UK Teaching Hospital* (Report). University of Leicester. Retrieved from <https://lra.le.ac.uk/handle/2381/39635>
- Toisoul, C. (2017). Design, Implementation and Evaluation of two Widgets for a MOOC's Learning Analytics Dashboard'. Retrieved from <http://dspace.ou.nl/handle/1820/7932>
- Tricot, A., Plégat-Soutjis, F., Camps, J.-F., Amiel, A., Lutz, G., & Morcillo, A. (2003). Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH (pp. 391–402). Presented at the Environnements Informatiques pour l'Apprentissage Humain 2003, ATIEF ; INRP. Retrieved from <https://edutice.archives-ouvertes.fr/edutice-00000154/document>