

Pour citer cet article : Roth, M. & Sánchez Abchi, V. (2020). Évaluer la difficulté d'une épreuve de compréhension de l'écrit. Expérimentation d'un outil sur une pratique enseignante. *La Revue LEEE*, 3. <https://doi.org/10.48325/rleee.003.02>

ÉVALUER LA DIFFICULTÉ D'UNE ÉPREUVE DE COMPRÉHENSION DE L'ÉCRIT

Expérimentation d'un outil sur une pratique enseignante

Murielle ROTH, Verónica SÁNCHEZ ABCHI

Version de la publication : septembre 2020

Évaluation ouverte et collaborative

Rétroacteur·rices : Fernando MORALES VILLABONA &
Patricia SCHILLINGS

Résumé

Cet article explore le problème de l'évaluation de la compréhension écrite dans la langue d'enseignement, en se concentrant notamment sur la régulation de la difficulté des instruments d'évaluation. Pour ce faire, nous nous intéressons à une pratique d'évaluation sommative de la compréhension de l'écrit d'une enseignante de 9^e (élèves âgés de 12-13 ans) mise en œuvre auprès de quatre classes d'élèves de trois niveaux scolaires différents. Nous analysons la difficulté du texte utilisé et des trois questionnaires - déclarés par l'enseignante comme étant de difficulté différente - afin de comprendre comment elle adapte les questions en fonction du niveau scolaire de ses élèves. Puis, nous comparons ces dimensions avec celles retenues dans un outil servant à évaluer la difficulté des textes et des questions en compréhension de l'écrit, développé par Sánchez Abchi et al. (2016). Les analyses montrent que les dimensions que l'enseignante fait varier pour ajuster la difficulté coïncident dans une large mesure avec les dimensions de l'outil d'estimation de la difficulté. Cependant, d'autres facteurs, liés au métier d'élève et aux pratiques d'enseignement, jouent également un rôle prépondérant dans l'évaluation de la compréhension écrite et dans l'estimation de la difficulté des tâches.

Mots-clés : épreuve, compréhension de l'écrit, difficulté du texte, difficulté des questions

Abstract

This article explores the problem of assessing reading comprehension in the school language, focusing in particular on regulating the difficulty of assessment instruments. To do so, we focus on a summative assessment practice of reading comprehension implemented by a 9th Grade teacher (pupils aged 12-13) in her class at three different levels. We analyze the difficulty of the text used and of the three questionnaires - declared by the teacher as being of different difficulty - in order to understand how she adapts the questions according to the grade level of her students. We then compare

these dimensions with those retained in a tool for evaluating the difficulty of texts and questions in reading comprehension, developed by Sánchez Abchi et al. (2016). The analyses show that the dimensions that the teacher varies to adjust the test difficulty coincide to a large extent with the dimensions in the tool for estimating difficulty. However, other factors, related to the pupils' habits and teaching practices, also play a major role in assessing reading comprehension and in estimating task difficulty.

Keywords: test, reading comprehension, difficulty of the text, difficulty of the questions

1. Introduction

Dans cet article, nous nous intéresserons à une pratique d'évaluation sommative de la compréhension de l'écrit mise en œuvre par une enseignante de 9^e (élèves âgé-es de 12-13 ans) avec ses élèves de trois niveaux scolaires différents. Nous analyserons la difficulté du texte utilisé et des trois questionnaires – déclarés par l'enseignante comme étant de difficulté différente – élaborés pour en évaluer la compréhension par les élèves des différents niveaux. Nous tenterons alors de comprendre comment elle adapte les questions en fonction du niveau scolaire de ses élèves, quelles dimensions de l'objet évalué – la compréhension d'un texte – sont amenées à être modifiées. Puis, nous comparerons ces dimensions avec celles retenues dans un outil servant à évaluer la difficulté des textes et des questions en compréhension de l'écrit, développé par Sánchez Abchi et al. (2016).

Aux activités d'apprentissage se mêlent des moments d'évaluation qui peuvent avoir une visée sommative cherchant à dresser un bilan des acquisitions des élèves (Cardinet, 1983 ; Perrenoud, 2001). Construire une épreuve sommative qui évalue la compréhension écrite des élèves demande aux enseignant-es des habiletés particulières pour que celle-ci soit valide, c'est-à-dire qu'il y ait une adéquation entre ce qu'on déclare évaluer et ce que l'on fait réellement (De Landsheere, 1988). Ces habiletés relèvent à la fois de méthodes propres à l'évaluation mais aussi des connaissances approfondies à propos de l'objet de savoir concerné. Il est notamment attendu de l'enseignant-e qu'il-elle détermine ses attentes, définisse des critères, s'interroge sur la manière dont l'information va être recueillie auprès des élèves mais aussi sur le matériel à lire – le texte utilisé et les questions associées – afin qu'il soit adapté au niveau des élèves et à l'objectif visé. La recherche a montré que si cette question de la difficulté du matériel à lire est une préoccupation bien présente chez les enseignant-es, l'estimation de son niveau de difficulté est souvent le résultat d'un examen subjectif (Mesnager, 2002) basé sur l'intuition ou l'expérience. C'est aussi le cas pour les questions qui sont généralement élaborées sans critères explicitement posés (Laveault & Grégoire, 2014).

La problématique de la difficulté des textes et des questions associées n'est pas nouvelle. En effet, les premiers travaux sur la lisibilité des textes datent des années 1920 et ont donné lieu à de nombreuses formules mathématiques permettant de mesurer la difficulté des textes, tout d'abord pour l'anglais (Flesch, 1948 ; Gunning, 1952) puis le français (De Landsheere, 1963; Kandel & Moles, 1958). Cependant, ces formules ont parfois été considérées comme insuffisantes, car elles ne prenaient en compte qu'une partie des spécificités du français (François, 2009). Quant aux questions qui interrogent le texte, elles ont également fait l'objet de plusieurs études qui ont montré qu'elles pouvaient être de difficulté variable en fonction de leur formulation (Zakhartchouk, 2000), du format de questionnement (Fletcher, 2006), des processus de lecture en jeu, si l'élève doit repérer, analyser, interpréter (OECD, 2019), etc. Dans les années 1990 et 2000, plusieurs recherches s'intéressant à l'évaluation de la compréhension de l'écrit se sont aussi interrogées sur le



choix des textes et la nature des questions mettant en évidence leur influence en termes de difficulté (Aeby et al., 2000 ; Broi & Wirthner, 1992 ; Weiss & Wirthner, 1991). Plus récemment, pour le français, Sánchez Abchi, et al. (2016) ont mené une étude exploratoire qui a conduit à créer un outil et une procédure pour évaluer *a priori* la difficulté des textes et des questions en compréhension de l'écrit.

Cette contribution est une première étape - d'un projet de thèse - qui vise à confronter l'outil d'estimation de la difficulté, créé par Sánchez Abchi et al. (2016) sur la base de l'analyse de tâches provenant d'épreuves cantonales de Suisse romande (épreuves externes), à des données empiriques d'une enseignante (épreuves internes). En d'autres termes, il s'agit de voir si l'outil correspond aux manières de faire de l'enseignante. Nous commencerons par analyser - à l'aide de cet outil - la difficulté d'un texte et de trois questionnaires (chacun composé de plusieurs questions qui interrogent le même texte) réalisés par une enseignante de cycle 3 (élèves âgés de 12-15 ans). Ceux-ci ont été conçus pour évaluer - de manière sommative - la compréhension de l'écrit d'élèves de 9^e (âgés de 12-13 ans) scolarisés dans trois filières différentes en fonction de leur niveau scolaire : filière à exigences élevées (A), filière à exigences moyennes (B) et filière à exigences élémentaires (C). Ensuite, nous effectuerons une comparaison entre les éléments intégrés de l'outil et les dimensions retenues par l'enseignante pour faire varier la difficulté.

L'objectif de cet article est de mieux comprendre les dimensions de difficulté que l'enseignante fait varier et dans quelle mesure celles-ci sont cohérentes avec l'outil de Sánchez Abchi et al. (2016), développé sur la base de l'analyse d'épreuves cantonales. Cette comparaison, qui est une première étape de recherche, devrait nous permettre de tester l'outil dans la perspective d'apporter d'éventuels ajustements et in fine de le mettre à disposition des enseignant·es. Plus spécifiquement, nous souhaitons répondre aux deux questions de recherche suivantes : 1) quelle est la difficulté des questions posées par l'enseignante dans les trois questionnaires ? ; 2) les dimensions de difficulté faisant l'objet de variation d'un questionnaire à l'autre par l'enseignante s'apparentent-elles à celles de l'outil mis à l'épreuve ou différent-elles ?

2. Le cadre de référence

2.1 La difficulté des textes et des questions

La compréhension de l'écrit telle qu'elle se manifeste dans l'évaluation est le résultat de l'interaction entre un lecteur, le matériel lu - un texte et [des questions qui rendent visible la compréhension] - et un contexte (Giasson, 2016), c'est pourquoi les difficultés auxquelles les élèves peuvent se heurter lors de la compréhension d'un texte et de ses questions ne sont jamais le fruit que d'un seul facteur. Cependant, de nombreuses études portant sur les processus d'apprentissage et sur l'évaluation de la compréhension écrite ont relevé l'importance de se préoccuper des caractéristiques des textes utilisés pour évaluer les performances des élèves (Lafontaine, 2003 ; Mesmer et al., 2012).

La recherche a bien exploré les éléments qui rendent un texte *difficile* en soi : la fréquence et le niveau de familiarité et d'abstraction des mots employés (Beck et al., 1982 ; Mesnager, 2002), les marques de cohésion utilisées, l'organisation de l'information, la familiarité de la thématique (Rawson & Kintsch, 2002), la structure du texte (Caccamise et al., 2013 ;



Mosenthal & Kirsch, 1998) et le genre textuel (Best et al., 2008), ainsi que l'articulation et les points de contact entre ces différents critères (Entin & Klar, 1985, cité par Boyer, 1992; Hiebert & Pearson, 2010). Cette liste n'est pas exhaustive, mais illustre certaines des dimensions qui peuvent avoir un impact sur le niveau de difficulté du texte.

Le texte a donc une certaine importance mais les questions adressées aux élèves pour rendre visible leur compréhension jouent également un rôle dans l'estimation de la difficulté de la tâche évaluative. La compréhension écrite constitue un processus complexe qui fait appel à des opérations cognitives diverses mettant en évidence différents aspects de ce processus. En effet, la lecture d'un texte suppose la mise en œuvre de processus de niveau inférieur, comme la maîtrise des correspondances entre les phonèmes et les graphèmes, et de processus de niveau plus élevé, comme la réalisation d'inférences de cause ou l'interprétation du sens *caché* d'un texte (pour une synthèse, voir Bianco, 2015). Si toutes les opérations sont nécessaires pour réussir à comprendre un texte, cela ne veut pas dire pour autant que toutes les tâches d'évaluation visent les mêmes opérations et les mêmes aspects du processus. En fonction du contexte dans lequel l'évaluation est réalisée (degré scolaire ; type d'élèves ; but de l'évaluation ; manière dont l'information va être recueillie auprès des élèves), chaque épreuve définit finalement les aspects de la compétence de compréhension qu'elle va focaliser et rendre visibles. Puis, une fois ces éléments définis, il s'agit d'élaborer des questions qui évaluent précisément ce qu'on a décidé d'évaluer - au niveau d'exigence qu'on attend des élèves.

La relation entre les questions de compréhension de l'écrit et les caractéristiques des textes a été comparativement moins exploitée (Kobayashi, 2004). Toutefois, certain·es auteurs et autrices suggèrent un lien entre l'objectif et/ou le type de(s) questions et les caractéristiques du texte (son genre). Par exemple, la recherche menée par Eason et al. (2012), en anglais, a montré que les questions littérales sont mieux réussies que les questions d'interprétation dans les textes narratifs, alors que les questions d'interprétation ont été mieux réussies que les questions littérales pour les textes expositifs. Ces résultats peuvent notamment s'expliquer par le fait que les textes expositifs et les questions complexes demandent davantage de capacités cognitives de haut niveau comme des capacités inférentielles et de raisonnement. De même, De Mier et al. (2013), dans une étude sur les processus de compréhension des enfants hispanophones, ont observé une performance différenciée selon le genre de texte - narratif versus expositif - et le niveau de compréhension que la question exige - littéral versus inférentiel. En général, les textes narratifs sont mieux compris et les questions littérales sont mieux réussies dans ces textes-là que dans les textes expositifs.

Nos propres recherches nous ont aussi laissé penser qu'en fonction du genre de texte mais surtout de ses caractéristiques, nous ne pouvons pas élaborer les mêmes types de questions. Par exemple, un *texte qui raconte* qui est très explicite et a une structure canonique ne permet pas de poser les mêmes questions qu'un texte qui a beaucoup d'implicite et des *flashbacks* (Roth et al., 2014). Si le texte influence effectivement le niveau de difficulté des questions, il peut tout de même être relativisé en fonction des pratiques sociales de la classe (Mottier Lopez, 2008).

2.2 Un outil pour évaluer la difficulté

Sánchez Abchi et al. (2016) ont créé un outil composé de trois indices de difficulté - un pour le matériel langagier (texte ou corpus), deux pour les questions (un pour le contenu et un pour l'enveloppe) - et une procédure qui ont pour objectif de guider les évaluateur·rices



dans l'analyse et l'évaluation de la difficulté des textes et des questions et de les aider à les choisir ou les modifier par rapport à la difficulté souhaitée. Chaque indice est lui-même constitué de plusieurs facteurs. La procédure détaillée du calcul de la difficulté du texte et des questions est présentée dans la section méthodologie. L'ensemble des facteurs définit en quelque sorte la manière dont la compréhension est appréhendée en relation aux caractéristiques du texte et permet de mieux saisir les difficultés potentielles que les élèves pourraient rencontrer (Roth et al., 2018 ; Sánchez Abchi et al., 2016).

2.1.1 *Le texte*

Pour estimer la difficulté du texte, la procédure comprend trois facteurs sélectionnés parmi ceux rencontrés dans nos lectures (François, 2009 ; Kandel & Moles, 1958 ; Lafontaine, 2003 ; McNamara et al., 2012) et réinterprétés sur la base d'un certain nombre d'analyses d'épreuves cantonales provenant de la Suisse romande (Roth et al., 2018) :

- La complexité lexicale - elle est déterminée par le pourcentage de mots peu fréquents, obtenu au moyen du logiciel de Mesnager et Bres (2008). Jusqu'à 5% de mots peu fréquents, un texte est censé être facilement accessible pour les élèves concernés, et à partir de 10% de mots peu fréquents, il est considéré comme difficile.
- La complexité syntaxique - elle est mesurée à l'aide de l'indice de Gunning (1952), qui par une formule mathématique, prend en compte le nombre total de mots, le nombre de phrases et le nombre de mots de trois syllabes et plus. La valeur obtenue se situe entre six et huit pour un texte simple et dès 12 pour un texte difficile.
- La complexité de la structure textuelle - elle est évaluée tout d'abord en définissant si celle-ci est typique du genre dont relève le texte. Le résultat est ensuite nuancé par la présence ou l'absence d'organiseurs textuels et leur pertinence.

L'indice de difficulté du texte obtenu à partir des trois grands facteurs retenus ci-dessus peut encore être nuancé par plusieurs critères qualitatifs sous forme de questionnement qui sont susceptibles d'apporter des « bonus » ou des « malus » (Sánchez Abchi et al., 2016).

2.1.2 *Les questions*

Dans cet article, nous nous intéresserons plus particulièrement aux questions étant donné que ce sont elles qui font l'objet de variation d'un questionnaire à l'autre. La difficulté de chacune d'elles est estimée sur la base de deux indices : le contenu et l'enveloppe. Chaque indice est lui-même composé de plusieurs dimensions.

La difficulté du contenu de la question est estimée sur la base :

- des opérations cognitives à mobiliser par l'élève ;
- de l'ampleur de l'objet langagier (renvoie à l'étendue de ce que l'élève doit avoir compris pour répondre à la question, le texte en entier, un paragraphe, une phrase ou un mot) ;
- et du nombre d'activités qu'il doit réaliser (concerne le nombre de réponses à donner).

La difficulté de l'enveloppe de la question est estimée en fonction :

- de la formulation de la consigne ;



- du format de questionnement et ce qu'il demande en plus comme travail pour répondre (questions à choix multiples, questions ouvertes demandant une activité d'écriture, etc.) ;
- et du matériel complémentaire mis à disposition de l'élève (dictionnaires, tableaux de conjugaison, etc.).

La difficulté du contenu de la question

Le processus de compréhension peut être caractérisé par trois niveaux : 1) le niveau de surface dans lequel il s'agit de comprendre les mots et les phrases ; 2) le niveau de base qui consiste à comprendre les propositions de manière individuelle mais aussi les liens entre elles dans le texte et en interaction avec les connaissances préalables du lecteur ; 3) le modèle de situation où l'élève devrait apprendre du texte grâce à une représentation de celui-ci qui est intégrée avec d'autres connaissances (préalables du lecteur) (Kintsch, 1988 ; McNamara & Kintsch, 1996). Ces niveaux de compréhension impliquent différents processus de lecture qui renvoient aux opérations cognitives que le lecteur et la lectrice mobilisent pour répondre à la question. Celles-ci peuvent être de difficulté variable. En principe, une question de repérage d'une information clairement formulée dans le texte (un mot ou une phrase) devrait être plus facile qu'une inférence référentielle ou la mobilisation d'une connaissance linguistique, telle qu'identifier l'antécédent d'un pronom. De même, une question qui exige une inférence sur les intentions ou les émotions de l'auteur·rice (Graesser et al., 1994) ou encore une question d'interprétation sont considérées comme plus exigeantes que les opérations évoquées précédemment. Par ailleurs, si la question implique d'effectuer non pas une, mais plusieurs opérations, la difficulté sera également plus grande. Cependant, l'ampleur de l'objet langagier est aussi un élément à prendre en compte dans la définition du niveau de difficulté du contenu de la question. En effet, si l'opération de repérage concerne un mot, la difficulté sera moindre que si elle concerne un paragraphe ou même qu'elle demande de parcourir l'ensemble du texte pour identifier plusieurs éléments. Et enfin, le nombre d'activités effectuées par l'élève est également considéré, car plus les activités sont nombreuses plus cela rend la tâche difficile. Par exemple, la question suivante « Relève quatre actions qui montrent la gentillesse de la sorcière » est plus difficile que si la demande avait été de relever une action.

La difficulté de l'enveloppe de la question

Concernant l'enveloppe de la question, les dimensions qui jouent un rôle dans l'estimation de la difficulté sont la formulation de la consigne et le format de questionnement. Une consigne formulée avec une négation, une syntaxe complexe et un vocabulaire difficile augmentera la difficulté de la question. Quant au format de questionnement, nous avons pris en compte les formats les plus courants dans les évaluations de la compréhension (Pini et al., 2006). Ensuite, nous les avons classés en fonction de la quantité d'activités qu'ils exigent. Par exemple, une *question à choix multiples* (QCM) avec quatre options dont une seule est juste donne un format plus simple qu'une *question d'ordination* pour laquelle les élèves sont appelés à sérier les éléments d'une liste, selon un critère donné (l'argument le plus fort, les événements d'une histoire d'après l'ordre chronologique, etc.). De même, une *question à réponse textuelle* où l'élève doit rédiger une réponse plus ou moins longue est plus complexe qu'une *question à réponses brève* pour laquelle l'élève doit produire un mot, une phrase ou un nombre. La combinaison de plusieurs formats dans une même question ajoute également de la difficulté à la tâche. Enfin, nous avons tenu compte de l'aide potentielle que certaines ressources peuvent apporter à l'élève qui passe l'épreuve. Celles-ci peuvent être un dictionnaire, un outil de conjugaison ou un ouvrage de référence.



Cependant, même si ces ressources sont disponibles, elles ne vont pas automatiquement faciliter la compréhension. Elles ne seront considérées comme un élément facilitateur que si elles peuvent clairement contribuer à la résolution de la question. Par exemple, le dictionnaire aidera à la compréhension si la question implique la mobilisation d'un vocabulaire.

3. Méthodologie

3.1 La pratique d'évaluation d'une enseignante

Nous avons collaboré avec une enseignante de français du cycle 3 (niveau secondaire I) du canton de Berne francophone, en Suisse. Il s'agit d'une enseignante expérimentée qui nous a fait part de ses réflexions sur ses pratiques évaluatives et qui nous a ainsi mis à disposition, de manière volontaire, une évaluation en compréhension de l'écrit élaborée pour ses élèves. Cette enseignante a des élèves de 9^e (âgé-es de 12-13 ans) répartis dans trois filières de formation. Celles-ci ont des niveaux d'exigences différentes : section préparant aux écoles de maturité (A : exigences élevées), section moderne (B : exigences moyennes) et section générale (C : exigences élémentaires). Dans ce contexte-là, lorsqu'elle élabore une épreuve, elle en fait donc plusieurs variantes en fonction du niveau de ses élèves.

A la fin d'une séquence d'enseignement sur la compréhension d'un texte qui transmet des savoirs, l'enseignante a décidé d'évaluer ses élèves de manière sommative pour faire le bilan de leurs acquisitions. Elle a donc choisi un article de journal qui traite du réchauffement climatique et a élaboré trois questionnaires - qu'elle considère comme étant de difficulté différente - à partir de la lecture de ce même texte.

3.2 Corpus

Les documents analysés se composent d'un texte, de trois questionnaires élaborés à partir de ce même texte, ainsi que d'un corrigé. Nous avons également obtenu les copies des questionnaires passés par 57 élèves (13 copies de la section A, 28 de la section B et 16 de la section C), avec les corrections effectuées.

Le texte

Le texte est un article de journal intitulé *Ça chauffe ! Notre climat en sursis* (Dürrenmatt, 2018) ([annexe A](#)). Il appartient au regroupement de genre *qui transmet des savoirs* (Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP), 2010) et a pour but d'informer les lecteur-rices sur les conséquences du réchauffement climatique. Après une introduction qui présente le thème, la suite du texte est principalement composée de séquences explicatives dans lesquels des séquences descriptives sont intercalées. Nous retrouvons en différents endroits du discours rapporté qui provient de spécialistes. Une autre particularité de ce genre de texte vient de la présence de marques graphiques telles que des illustrations et des photos (Chartrand, 2015).



Les questionnaires

Le questionnaire étant le genre scolaire privilégié pour amener les élèves à réfléchir à un texte (Habi, 2016), l'enseignante a choisi cette modalité pour évaluer leur compréhension de l'écrit dans les trois sections (A, B, C). Le questionnaire A comprend 19 questions ([annexe B](#)), le B 18 questions ([annexe C](#)), et le C 14 questions ([annexe D](#)). Les questions se distribuent de la manière suivante : certaines sont communes à deux ou trois des questionnaires et d'autres sont spécifiques à un des questionnaires (voir tableau 1).

Tableau 1

Nombre de questions présentes dans les différents questionnaires

	Nombre de questions	Questionnaires
Questions spécifiques à un questionnaire	7	A
	1	B
	2	C
Questions communes	7	A - B - C
	5	A - B
	5	B - C

Voici un exemple de question avec ses variantes, c'est-à-dire qui se décline différemment dans les trois questionnaires :

- Question 10A : Quelle est la conséquence de la fonte des glaciers pour la biodiversité ?
- Question 10B : Quelle est la conséquence de la fonte des glaciers pour certains animaux ?
- Question 10C : Quelle est la conséquence de la fonte des glaciers pour certains animaux ?

- Ils auront plus chaud.
- Ils doivent aller plus haut pour trouver de la nourriture.
- Ils doivent aller plus bas pour trouver de la nourriture.

3.3 Démarche

Pour répondre à notre première question de recherche - quelle est la difficulté des questions posées par l'enseignante dans les trois questionnaires ? -, nous avons analysé - à l'aide de l'outil de Sánchez Abchi et al. (2016) - la difficulté du texte et de l'ensemble des questions provenant des trois questionnaires. Autrement dit, nous avons examiné les différentes dimensions retenues par l'outil d'estimation : les opérations cognitives censées être mobilisées par l'élève pour répondre aux différentes questions, l'ampleur de l'objet langagier, le nombre d'activités à réaliser par l'élève, la formulation de la consigne et le format de questionnement. Par ailleurs, les élèves n'avaient pas accès à du matériel complémentaire tel qu'un dictionnaire ou d'autres ouvrages de référence pendant l'évaluation. Nous avons également estimé le niveau de difficulté du texte à comprendre (Dürrenmatt, 2018) compte tenu de la difficulté de son vocabulaire, de son niveau de complexité syntaxique et de la complexité de sa structure.

Pour répondre à notre deuxième question de recherche - les dimensions de difficulté faisant l'objet de variation d'un questionnaire à l'autre par l'enseignante s'apparentent-elles



à celles de l'outil mis à l'épreuve ou différentes ? –, nous avons comparé ces dimensions avec celles qui ont été effectivement modifiées par l'enseignante dans les trois questionnaires. Ainsi, nous devrions saisir dans quelle mesure les dimensions de l'outil rejoignent celles que l'enseignante prend en considération pour faire varier la difficulté des questions.

Quant aux résultats des élèves et aux corrections, nous ne les avons pas analysés de manière spécifique mais nous avons utilisé ces informations pour mieux caractériser nos données : les réponses attendues sont explicitées à l'aide des corrections réalisées dans les questionnaires.

3.4 La procédure de calcul de la difficulté

Pour saisir la difficulté d'une épreuve en compréhension d'un texte, nous nous sommes basés sur l'outil présenté en détails par Sánchez Abchi et al. (2016). Trois indices juxtaposés sont nécessaires – un pour le texte et deux pour les questions. Ces indices devraient pouvoir guider les évaluateur·rices dans l'analyse et l'estimation de la difficulté (voir tableau 2).

Tableau 2

Outil d'estimation de la difficulté du texte et des questions (Sánchez Abchi et al., 2016)

Indice Texte			Indice Contenu questions			Indice Enveloppe questions		
Vocabulaire	Complexité syntaxique	Structure texte	Type opération cognitive	Ampleur objet langagier	Nombre activité à réaliser	Consigne	Format de questionnaire	Matériel à disposition
Va	Vb	Vc	Vd	Ve	Vf	Vg	Vh	Vi
= $Va + Vb + Vc / 3$			= $Vd + Ve + Vf / 3$			= $Vg + Vh + Vi / 3$		
Valeur moyenne Texte convertie en indice qualitatif en se référant au Tableau 1 puis pondération par bonus et malus			Valeur moyenne Contenu questions convertie en indice qualitatif en se référant au Tableau 1			Valeur moyenne Enveloppe questions convertie en indice qualitatif en se référant au Tableau 1		
Indice qualitatif Texte			Indice qualitatif Contenu questions			Indice qualitatif Enveloppe questions		

Afin de prendre en compte les trois indices (texte, contenu et enveloppe), l'outil permet de calculer ou d'estimer la difficulté de chacune des dimensions qu'ils contiennent, sur une échelle comportant trois valeurs : -1 (= facile), 0 (= intermédiaire) et 1 (= difficile). Ensuite, pour chaque indice, une moyenne des valeurs est établie puis mise en relation avec un tableau sous forme d'échelle nominale à cinq niveaux. Puis, la valeur obtenue pour chaque indice est exprimée sous forme qualitative, chacun des indices ayant le même poids (voir tableau 3).



Tableau 3

Echelle nominale à cinq niveaux pour déterminer l'indice de difficulté du texte et des questions associées (Sánchez Abchi et al., 2016)

Valeurs	Indice de difficulté qualitatif
-1 à -0.61	Facile (F)
-0.6 à -0.21	Plutôt facile (PF)
-0.2 à 0.2	Intermédiaire (I)
0.21 à 0.6	Plutôt difficile (PD)
0.61 à 1	Difficile (D)

Ainsi, par exemple, la question « A quelle date a-t-il [été] écrit ? Par qui ? » demande à l'élève de *repérer* des éléments (opération cognitive), il a *deux activités* à mener (repérage de la date et de l'auteur-riche) et l'ampleur de l'objet langagier est la *phrase* (les deux informations se trouvent dans une courte phrase). Concernant les dimensions de l'enveloppe, la consigne comprend un *vocabulaire et une syntaxe simples* et le format de questionnement est qualifié de *question à réponse brève* (date et nom).

Ensuite, pour chaque dimension, une des valeurs suivantes a été attribuée : -1 (facile), 0 (= intermédiaire) ou 1 (difficile). Puis, nous avons calculé la moyenne pour le contenu et pour l'enveloppe. Le tableau 4 illustre la procédure d'analyse détaillée.

Tableau 4

Exemple d'analyse détaillée d'une question présente dans les trois questionnaires

Question	Contenu			Enveloppe	
	Opérations cognitives	Nombre d'activités	Ampleur objet langagier	Consigne	Format de questionnement
A quelle date a-t-il [été] écrit ? Par qui ?	repérer	2	phrase	vocabulaire et syntaxe simples	question à réponse brève
	-1 (facile)	-1 (facile)	-1 (facile)	-1 (facile)	0 (intermédiaire)
	-1 = facile			-0.5 = plutôt facile	

4. Résultats

4.1 Analyse de la difficulté

4.1.1 Du texte

Pour le texte, qui appartient au genre qui transmet des savoirs, nous avons tout d'abord évalué - à l'aide de l'outil - de manière quantitative sa complexité lexicale, sa complexité syntaxique et la complexité de sa structure puis nous nous sommes interrogées sur l'éventualité de nuancer les premières mesures par des bonus ou malus, comme la procédure le permet. Avec plus de 10% de mots absents, le lexique est considéré comme difficile (valeur 1). Pour la syntaxe, la valeur obtenue étant de plus de 16, l'indice a été considéré comme difficile (valeur 1). Quant à la structure du texte, nous avons estimé qu'elle était facile (valeur -1), étant donné qu'elle est caractéristique du genre article de journal et que la faible présence d'organisateur textuels n'entrave pas la compréhension. Finalement, en faisant la moyenne des trois valeurs $((1+1+(-1))/3)$, nous obtenons pour l'indice texte une valeur de 0.33 qui correspond à un texte *plutôt difficile* (voir tableau 3).



Notons que l'enseignante travaille avec le même texte pour les trois niveaux d'exigences, donc ce n'est pas le texte lui-même qui est susceptible d'être modifié pour ajuster le niveau de difficulté. Dans la conception de l'enseignante, c'est fondamentalement les questions qui vont permettre de réguler la difficulté en fonction des différents niveaux de ses élèves. C'est justement sur cet aspect que nous allons nous focaliser maintenant.

4.1.2 Des questions pour les trois questionnaires

Pour chacune des questions, nous avons évalué la difficulté de son contenu (l'opération cognitive mobilisée, le nombre d'activités à réaliser et l'ampleur de l'objet langagier) et de son enveloppe (la consigne et le format de questionnement). Le tableau 5 présente la répartition des dimensions de difficulté dans les questions des trois questionnaires (A, B, C).

Tableau 5

Répartition des dimensions de difficulté dans les questions des trois questionnaires

Indices de difficulté	Dimensions de l'outil	Éléments des dimensions	Pourcentage de questions %		
			A	B	C
Contenu	Opération cognitives	repérer	37	50	57
		repérer + inférer	0	6	7
		inférer	58	38	29
		interpréter	5	6	7
	Nombre d'activités	1 activité	69	78	71
		2 activités	26	22	29
		3 activités	5	0	0
	Ampleur de l'objet langagier	mot	16	17	22
		phrase	37	44	50
		fragment	42	33	21
texte		5	6	7	
Enveloppe	Consigne	simple	63	72	71
		plus difficile	37	28	29
	Format de questionnement	QCM	0	11	29
		vrai-faux	0	6	7
		question à réponse brève	26	28	35
		question à réponse textuelle	74	55	29

Chaque questionnaire comporte un nombre différent de questions qui nous informe déjà sur une forme de régulation de la difficulté en lien avec la longueur de l'épreuve. D'un questionnaire à l'autre, nous observons des différences dans le contenu mais également dans certaines dimensions de l'enveloppe. Par rapport aux opérations cognitives, le type et la fréquence sont différents d'un questionnaire à l'autre. En effet, nous constatons une proportion plus grande de questions inférentielles dans le questionnaire dit *difficile* (58%) que dans le questionnaire le plus facile (29%). En ce qui concerne le nombre d'activités et l'ampleur de l'objet langagier, les différences entre les questionnaires sont moins importantes comme nous le verrons par la suite dans les analyses plus détaillées. Du point de vue des consignes, elles sont plus difficiles dans le questionnaire A que dans les deux autres questionnaires. Au sujet du format de questionnement, le questionnaire A comporte uniquement des questions ouvertes tandis que les questionnaires B et C contiennent à la fois des questions ouvertes et fermées avec une proportion deux fois plus grande de ces dernières pour le questionnaire C. Par ailleurs, le fait que plusieurs mêmes questions se retrouvent dans deux ou trois des questionnaires nous montre une certaine volonté de base



commune pour les trois niveaux d'exigences. Dans les sections suivantes nous nous concentrerons uniquement sur les questions communes et les questions avec variantes.

Les questions communes aux trois questionnaires

Nous avons poursuivi par l'analyse des sept questions communes aux trois questionnaires du point de vue de la difficulté de leur contenu et de leur enveloppe (voir tableau 6). Concernant les opérations cognitives, nous observons une distribution similaire entre les questions de repérage et celles qui requièrent d'effectuer des inférences simples à partir des idées ou informations contenues dans le texte. Une seule question exige de l'interprétation, c'est-à-dire de mettre en lien des informations locales du texte avec ses connaissances du monde. Pour la plupart des questions (cinq) l'élève doit réaliser une seule activité. Quant à l'ampleur de ce que l'élève doit avoir compris pour répondre à la question, nous remarquons une volonté de couvrir tous les niveaux de difficulté, car les réponses concernent aussi bien le mot que le texte en passant par la phrase et le fragment. Pour la consigne, il y a presque parité entre les consignes simples et celles qui sont plus difficiles. Finalement, pour le format de questionnement, nous retrouvons uniquement des questions à réponse brève ou à réponse textuelle, pas de vrai-faux ni de QCM. Le tableau 6 présente l'analyse des dimensions de la difficulté pour les questions communes. Les chiffres indiquent le nombre de questions concernées par chaque dimension.

Tableau 6

L'analyse des dimensions de la difficulté pour les sept questions communes aux trois questionnaires

Contenu des questions									
Opérations cognitives			Nombre d'activités		Ampleur de l'objet langagier				
repérer	inférer	interpréter	1 activité	2 activités	mot	phrase	fragment	texte	
3	3	1	5	2	2	3	1	1	
Enveloppe des questions									
Consigne				Format de questionnement					
vocabulaire et syntaxe simples		et vocabulaire spécifique et nombre de réponses attendues pas spécifié		1x pas		question à réponse brève		questions à réponse textuelle	
4		3		4		3			

Lorsque nous regardons la difficulté de chacune des sept questions, nous constatons - pour le contenu - que l'enseignante fait davantage varier la difficulté de l'opération cognitive que celle du nombre d'activités ou encore de l'ampleur de l'objet langagier. Toutefois, l'indice de difficulté du contenu étant le résultat d'une moyenne entre trois dimensions, toutes les questions sont considérées comme *faciles* sauf la dernière qui est *plutôt difficile*. Quant à l'enveloppe, elle est *plutôt facile* pour les trois premières questions, *intermédiaire* pour les deux questions suivantes et finalement *plutôt difficile* pour les deux dernières questions. Par ailleurs, lorsque nous observons le placement des questions tout au long de chacun des questionnaires, nous remarquons que les questions avec un contenu et une enveloppe plus difficiles se retrouvent vers la fin. C'est une manière pour l'enseignante de moduler la difficulté de manière progressive.

Les questions avec leur(s) variante(s)

Ce groupe de questions est composé de cinq questions de base qui ont une ou deux variantes réparties différemment dans les questionnaires. Le tableau 7 présente la distribution des dimensions de difficulté dans ce groupe de questions.



Tableau 7

Distribution des dimensions de la difficulté dans les questions et leur(s) variante(s)

Indices de difficulté	Dimensions de l'outil	Éléments des dimensions	Nombre de questions		
			A	B	C
Contenu	Opération cognitives	Repérer	3	3	3
		repérer + inférer	0	1	1
		Inférer	2	1	1
		Interpréter	0	0	0
	Nombre d'activités	1 activité	3	4	4
		2 activités	2	1	1
	Ampleur de l'objet langagier	Mot	1	1	1
		Phrase	3	4	4
		Fragment	1	0	0
		Texte	0	0	0
Enveloppe	Consigne	Simple	3	5	5
		plus difficile	2	0	0
	Format de questionnement	QCM	0	2	4
		question à réponse brève	1	1	1
		question à réponse textuelle	4	2	0

La modification des opérations cognitives d'un questionnaire à l'autre présente différents cas de figure : parfois, c'est la variante A qui demande une opération plus difficile (inférence) que pour les variantes B et C (repérage) ; parfois toutes les variantes demandent la même opération. Ainsi, cette dimension du contenu ne semble pas être déterminante. Nous constatons également qu'il y a relativement peu de variation entre les questionnaires concernant le nombre d'activités et l'ampleur de l'objet langagier. Pour la totalité des questions sauf une, le contenu est *facile* peu importe le niveau d'exigences (A, B ou C).

Par rapport à l'enveloppe, toutes les questions sauf deux comportent une consigne avec un vocabulaire et une syntaxe simples. Toutefois, nous observons que la formulation de la consigne devient moins précise dans le questionnaire difficile (A) où dominent les questions à réponse textuelle. En effet, lorsque nous considérons les corrections de l'enseignante, nous constatons que la bonne réponse, la réponse attendue de la part de l'élève, ne peut pas toujours être inférée à partir de la consigne.

Prenons par exemple la question « Et toi, que pourrais-tu faire concrètement pour réduire tes émissions de CO₂ ? », pour le questionnaire dit « difficile ». La consigne ne le dit pas, mais il semble qu'il était attendu que l'élève ne dise pas seulement ce qu'il pourrait faire mais comment il pourrait le faire. Les élèves qui n'ont donc pas répondu avec cette précision ont été pénalisés. Cet exemple montre que la formulation contribue à la difficulté de la question.

Quant au format de questionnement, nous trouvons une majorité de QCM dans le questionnaire facile tandis qu'il y a plus de questions à réponse textuelle dans le questionnaire dit difficile. Et c'est dans le questionnaire intermédiaire que nous constatons une distribution similaire de tous les formats. Le niveau de difficulté de l'enveloppe varie davantage que pour le contenu.

En résumé, les niveaux de difficulté *facile* et *plutôt facile* se rapportent aux niveaux d'exigences B et C, alors que les niveaux de difficulté *intermédiaire* et *plutôt difficile* aux niveaux d'exigences A et B. Ainsi, il apparaît, comme pour les questions communes, que c'est la difficulté de l'enveloppe qui fait l'objet de variations et pas le contenu.



4.2 Comparaison entre les dimensions de l'outil et celles modifiées par l'enseignante

Les analyses réalisées (voir chapitre 4.1.2) nous permettent maintenant de comparer chacune des dimensions de l'outil avec celles que l'enseignante fait varier et voir dans quelle mesure il y a une correspondance ou non entre les deux. Concrètement, pour effectuer cette comparaison, nous avons pris chacune des dimensions et avons observé si l'enseignante opérait effectivement une modification de celles-ci d'un niveau d'exigence à l'autre.

Pour les questions présentes dans les trois questionnaires qui constituent en quelque sorte la base commune de ce qui est attendu pour les trois niveaux d'exigences, la dimension « opération cognitive » est prise en compte par l'enseignante dans une perspective de progression à l'intérieur de l'épreuve. La difficulté va crescendo avec tout d'abord des questions de repérage, ensuite d'inférences et finalement une interprétation. Le nombre d'activités n'est pas une dimension que l'enseignante fait beaucoup varier. Parfois le questionnaire A comporte une activité de plus à faire que pour les questionnaires B et C. L'ampleur de l'objet langagier est également une dimension prise en compte par l'enseignante, en particulier en ce qui concerne les questions communes aux trois questionnaires. Elle devient plus importante - mot, phrase, fragment, texte - au fur et à mesure de l'avancée dans les questionnaires.

La difficulté du contenu est modulée à l'intérieur des épreuves pour les questions communes en particulier la dimension opération cognitive. Toutefois, ce n'est pas le cas pour les questions avec des variantes entre les questionnaires. En effet, comme observé dans la section d'avant, pour moduler le niveau de difficulté entre les questionnaires, ce n'est pas le contenu qui est modifié davantage, mais c'est surtout sur l'enveloppe que l'enseignante se concentre. Effectivement, le format de questionnement est une dimension que l'enseignante fait varier. Ainsi, celui-ci est simplifié pour la variante C et parfois B par rapport à la variante A. La consigne est un élément pris en compte par l'enseignante autant pour les questions communes que pour les questions avec variantes.

Pour les questions avec variantes, toutefois, il faut signaler ici que, au-delà de quelques termes plus difficiles à comprendre comme par exemple *rubrique*, *biodiversité*, *surface totale*, les complications sont dues à un certain manque de précision dans la variante difficile de la question, qui la rend parfois ambiguë. Dès lors, cela pourrait donner lieu à plusieurs réponses possibles alors même qu'une seule réponse est attendue de la part des élèves (d'après le corrigé fourni par l'enseignante et la manière dont les copies des élèves ont été corrigées). La consigne n'est pas totalement explicite. Ainsi, ce qui détermine le niveau de difficulté est la capacité des élèves à *deviner* ce qui est attendu de leur part. Pour les élèves qui sont évalués avec les questionnaires B et C, les questions sont, par contre, moins ambiguës.

Finalement, une dernière dimension prise en compte par l'enseignante ne fait pas partie de celles retenues par l'outil : la longueur du questionnaire, c'est-à-dire le nombre de questions présentes. Ainsi, nous constatons que les questions sont plus nombreuses pour les élèves de niveau A (19) que pour les élèves de niveau C (14) avec le même temps mis à disposition pour faire l'épreuve. La longueur n'a pas été considérée dans l'outil puisque celui-ci a été créé dans une logique de tâches et non d'épreuve comme c'est le cas ici.



5. Discussion

Dans cet article, nous nous sommes intéressées à la problématique de la difficulté de l'évaluation de la compréhension écrite en nous focalisant sur les pratiques d'évaluation d'une enseignante. Nous nous concentrons tout d'abord sur la difficulté des questions d'évaluation posées par l'enseignante dans les trois versions du questionnaire, sous l'angle de l'outil d'estimation de la difficulté élaboré par Sánchez Abchi et al. (2016). Ensuite nous réfléchissons aux implications des pratiques de régulation de la difficulté pour l'enseignement et finalement à la pertinence d'un tel outil pour soutenir les pratiques d'évaluation des enseignant-es à l'école.

L'enseignante a conçu trois épreuves de difficulté différente à partir de la lecture du même texte. Elle a adapté la difficulté de chacun des questionnaires en fonction du niveau d'exigence de la section. Le questionnaire A comporte une proportion plus importante de questions difficiles par rapport aux questionnaires B et C, notamment en raison des opérations cognitives – avec 63% d'inférence ou d'interprétations –, et du format de questionnement – avec presque 75% de questions à réponse textuelle. Même si l'enseignante régule la difficulté, d'un questionnaire à l'autre, en fonction du niveau de ses élèves, il existe un ensemble de questions communes à tous et toutes quel que soit leur niveau (A, B ou C). Ces questions constituent une palette assez large des possibilités évaluatives avec une certaine diversité qui entre en jeu. En effet, bien que cela ne soit pas dans les mêmes proportions, il y a une volonté de toucher à tous les niveaux de compréhension (repérer, inférer, interpréter), de faire varier le nombre d'activités, l'ampleur de ce qui est à comprendre, la difficulté des consignes et les types de formats de questionnement. Nous pouvons supposer que cet éventail est en quelque sorte les exigences minimums qu'elle attend de n'importe quel élève de 12-13 ans en compréhension de cet article de journal. De plus, l'enseignante a aussi réfléchi au placement de ces questions communes, tout au long de chaque questionnaire, dans une logique de progression du plus facile au plus difficile qui concerne avant tout l'enveloppe et non pas le contenu.

À côté des questions communes, l'enseignante a élaboré cinq questions pour lesquelles elle a prévu une ou deux variantes en fonction du niveau d'exigences (A, B ou C). Pour ces questions-là, l'éventail des modifications est moins grand que pour les questions communes, et c'est aussi la difficulté de l'enveloppe qui fait l'objet de modifications et pas le contenu. Pour ces cinq questions et leur(s) variante(s), nous supposons que l'enseignante a voulu réaliser un équilibre entre un texte plutôt difficile et des questions au contenu facile. Cette logique de simplification des questions a également été mise en évidence par Goigoux et Thomazet (1999) dans une étude sur les pratiques et conceptions de l'enseignement de la lecture.

De manière générale, les analyses nous montrent que les dimensions de l'outil pour estimer la difficulté des questions – les opérations cognitives, le nombre d'activités, l'ampleur de l'objet langagier, la consigne et le format de questionnement – sont aussi celles que l'enseignante fait varier d'un questionnaire à l'autre. Ainsi, la conception des dimensions de difficulté qui est à la base de l'outil rejoint en grande partie celle de l'enseignante. Toutefois, un certain nombre d'observations peuvent être réalisées et amènent à nuancer nos propos et à soulever quelques questionnements.

Tout d'abord, l'outil n'a pas de dimension qui permet de prendre en compte la difficulté liée à une logique d'épreuve qui demande de considérer la longueur du questionnaire – avec un nombre de questions auxquelles l'élève doit répondre dans un temps imparti – et



le positionnement des questions dans celui-ci en fonction de leur difficulté. Nous pouvons aussi relever une autre manière que l'enseignante a de réguler la difficulté en manipulant la dimension consigne. En effet, la formulation de certaines consignes, en particulier pour les questions ouvertes, nous laisse penser que l'enseignante est plus précise avec les élèves en difficulté, qui sont dans les sections à exigences élémentaires (C) et moyennes (B), qu'avec les élèves qui sont dans un niveau d'exigences élevées (A). Ce manque de précision demande à l'élève d'anticiper les attentes de l'enseignante pour pouvoir répondre correctement à la question. C'est donc moins la compréhension écrite de l'élève qui est évaluée que sa capacité à bien faire son *métier d'élève* (Perrenoud, 2017). Cette façon de procéder nous interroge, car dans certains cas plusieurs réponses nous paraissent possibles alors qu'une seule est considérée comme acceptable. Il y aurait donc inadéquation entre la demande qui émerge de la consigne et ce qui est visé, compromettant ainsi la validité didactique du questionnaire, du moins des questions concernées (De Pietro & Roth, 2017 ; Roth et al., 2019). Ce problème de validité, qui a été rendu visible par l'outil d'estimation de la difficulté, est lié, nous semble-t-il, au construit de la compréhension. Il nous est apparu en effet que la *profondeur* de compréhension du texte, telle qu'elle est évaluée ici, n'est pas la même dans les trois versions de l'épreuve. Plus les questions de l'épreuve sont faciles - avec beaucoup de repérage facile à corriger - plus la validité est forte. Il y a donc bien adéquation entre la demande et la réponse attendue mais nous restons à un niveau de compréhension de surface. Au contraire, plus les questions sont difficiles - avec davantage de questions ouvertes, qui demandent de faire des inférences, d'interpréter le texte, et qui sont donc plus compliquées à corriger - plus la validité devient problématique : la marge d'interprétation concernant ce qui est défini et accepté comme bonne réponse s'accroît en même temps qu'on vise un niveau de compréhension plus profond.

La régulation de la difficulté effectuée par l'enseignante pour mieux s'adapter aux niveaux des élèves fait émerger un autre paradoxe. En effet, pour éviter de mettre les élèves les plus faibles en échec, l'enseignante va éviter les questions qui touchent aux niveaux de compréhension les plus élevés et ainsi réduire les occasions que ceux-ci ont de progresser en compréhension. Ainsi, les ajustements réalisés de manière à ce que tous les élèves puissent réussir contribuent en même temps à renforcer les inégalités scolaires (Bonnéry, 2009 ; Joigneaux, 2009).

En ce qui concerne les possibilités d'exploitation de l'outil mis à l'épreuve ici, notons également que pour la difficulté du contenu des questions qui forment le socle commun, nous n'obtenons pratiquement que des indices de difficulté faciles, malgré le fait que l'enseignante fasse varier le type d'opération cognitive de manière à placer les questions plus difficiles à la fin des questionnaires. Ce résultat est dû au fait que l'outil fait la moyenne entre trois dimensions qui ont le même poids : l'opération cognitive à mobiliser, le nombre d'activités et l'ampleur de l'objet langagier. Or, même si la difficulté de l'opération cognitive à mobiliser est intermédiaire ou difficile, comme c'est le cas dans plusieurs de ces questions, elle est contrebalancée par les autres dimensions qui sont faciles. Dans cette situation, l'indice de difficulté du contenu ne reflète pas la difficulté réelle, posant la question du poids plus important qu'il faudrait peut-être donner à certaines dimensions par rapport aux autres.

Suite à ces quelques réflexions, nous pouvons nous interroger sur la pertinence de l'outil, c'est-à-dire sur sa capacité à donner une *bonne vision* des dimensions du texte et des questions et de leur difficulté. De manière générale, il nous semble que l'outil offre un panorama des dimensions en jeu dans le texte et les questions et qui pourraient potentiellement être difficiles pour les élèves. Cependant, un certain nombre de constats



nous suggèrent d'affiner l'outil notamment en le confrontant aux pratiques de classe. Nous faisons l'hypothèse que celles-ci vont jouer un rôle lorsque l'enseignant·e évalue la difficulté du texte et des questions associées. L'enseignant·e qui habitue ses élèves à certaines pratiques d'évaluation, en particulier concernant la forme des questions, aura probablement une autre perception de la difficulté de celles-ci pour ces élèves. De même, si les élèves ont une certaine familiarité avec le contenu thématique du texte à lire (Denton et al., 2015 ; Kobayashi, 2009). Par exemple, un texte qui porte sur les esquimaux dans le grand nord pourrait être considéré comme difficile par l'enseignant·e, car la thématique est éloignée du contexte d'élèves de 8^e. Mais ce même texte pourrait aussi être considéré comme facile pour des élèves qui auraient réalisé tout un travail autour de cette même thématique. Et finalement, nous supposons qu'une meilleure connaissance des pratiques d'enseignement-apprentissage nous permettrait ensuite de nuancer la difficulté de ce qui est évalué. Par exemple, les questions qui portent sur la connaissance des éléments spécifiques au genre de texte évalué peuvent être considérées comme faciles ou difficiles en fonction de l'enseignement qui aurait été réalisé en classe.

6. En guise de conclusion

Cette contribution se voulait être une première étape dans un processus de confrontation entre un outil développé sur une base théorique et au contact de l'analyse d'épreuves cantonales de Suisse romande, et des données empiriques provenant des pratiques évaluatives d'une enseignante. Cette étude nous montre que l'outil permet d'estimer la difficulté du texte et des trois questionnaires utilisés pour évaluer la compréhension écrite d'élèves de 9^e. De plus, les dimensions retenues par l'outil coïncident largement avec la conception que l'enseignante se fait des dimensions qui doivent varier en difficulté. Cette expérimentation nous montre aussi que si les dimensions de difficulté rejoignent globalement celle de l'enseignante, il est nécessaire d'affiner la compréhension de chacune d'elles - par exemple, la conception de ce que c'est une consigne difficile -, et de penser à leur poids dans le calcul de la difficulté en particulier des questions.

Les conclusions de cette étude, qui a été réalisée à partir de documents fournis par une enseignante, doivent être interprétées avec prudence et ne peuvent prétendre à être généralisées. Néanmoins, ces premiers résultats confirment la pertinence de cet outil d'estimation de la difficulté en particulier des questions. Des études complémentaires devraient être mises en œuvre - en tenant compte des pratiques de classe - pour pouvoir affiner encore l'outil de manière à ce qu'il puisse être un guide fiable pour les enseignant·es dans la caractérisation des dimensions du texte et des questions associées qui déterminent leur difficulté lorsqu'ils sont exploités pour évaluer la compréhension écrite d'élèves de 10 à 15 ans. Comme nos données permettent de le supposer, les contextes devraient avoir un impact sur l'évaluation de la difficulté des textes et des questions utilisés pour l'évaluation. Dans cette perspective, il pourrait être envisageable de travailler de manière collaborative avec un petit groupe d'enseignant·es de manière à mettre en discussion l'outil, le transformer ou encore d'en élaborer un nouveau. Cette volonté d'approcher la classe et plus particulièrement les enseignant·es vient d'une prise au sérieux de la perspective située des apprentissages (Mottier Lopez, 2008). De plus, pour s'assurer que l'outil est un bon prédicteur de difficulté, il serait nécessaire d'analyser dans quelle mesure un lien est possible entre l'analyse a priori de la difficulté des textes et des questions - réalisée à l'aide de l'outil - et la réussite des élèves.



Par ailleurs, cette étude nous a permis d'attirer l'attention sur l'importance d'avoir une vision claire des objets mobilisés pour évaluer les élèves en compréhension écrite et la difficulté que ceux-ci peuvent potentiellement comporter. Faute de quoi, il pourrait y avoir un problème de validité de contenu (De Pietro & Roth, 2016) lorsqu'il y a non-adéquation entre le niveau de difficulté des textes et des questions et les compétences et connaissances de l'élève que l'on cherche à évaluer.

Références

- Aeby, S., De Pietro, J.-F., & Wirthner, M. (2000). *Français 2000 : dossier préparatoire. L'enseignement du français en Suisse romande : un état des lieux et des questions*. Institut de recherche et de documentation pédagogique.
- Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74(4), 506-521. <https://doi.org/10.1037/0022-0663.74.4.506>
- Best, R. M., Floyd, R. G., & Mcnamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*, 29(2), 137-164. <https://doi.org/10.1080/02702710801963951>
- Bianco, M. (2015). *Du langage oral à la compréhension de l'écrit*. Presses universitaires de Grenoble.
- Bonnéry, S. (2009). Scénarisation des dispositifs pédagogiques et inégalités d'apprentissage. *Revue française de pédagogie*, 167, 13-23. <https://doi.org/10.4000/rfp.1246>
- Boyer, J.-Y. (1992). La lisibilité. *Revue française de pédagogie*, 99(1), 5-14. <https://doi.org/10.3406/rfp.1992.1322>
- Broi, A.-M., & Wirthner, M. (1992). *Lorsque le Mississippi passe par Neuchâtel et Fribourg : évaluation de la compréhension de la lecture en 4P : résumé*. Institut romand de recherches et de documentation pédagogiques.
- Caccamise, D., Friend, A., Kintsch, W., & Kintsch, E. (2013, July 16-18). *Measuring text complexity*. [Poster presentation]. The Twenty-third annual meeting of the society for text & discourse, Valencia, Spain.
- Cardinet, J. (1983). *Des instruments d'évaluation pour chaque fonction*. Institut romand de recherches et de documentation pédagogiques.
- Chartrand, S.-G. (2015). *Caractéristiques de 50 genres pour développer les compétences langagières en français (2^e éd.)*. Didactica.
- Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP). (2010). *Plan d'études romand (PER)*. CIIP. <http://www.plandetudes.ch/>
- De Landsheere, G. (1963). Pour une application des tests de lisibilité de Flesch à la langue française. *Le Travail Humain*, 26(1/2), 141-154.
- De Landsheere, V. (1988). *Faire réussir, faire échouer*. Presses universitaires de France.
- De Mier, M. V., Borzone, A. M., Sánchez Abchi, V., & Benítez, M. E. (2013). Habilidades de comprensión y factores textuales en los primeros grados. *Revista de Peruana de Psicología y Trabajo Social*, 2(1), 89-106.
- De Pietro, J.-F., & Roth, M. (2017). A propos de la validité « didactique » d'une évaluation. *Evaluer : journal international de recherche en éducation et formation*, 3(3), 31-50.
- Denton, C. A., Enos, M., York, M. J., Francis, D. J., Barnes, M. A., Kulesz, P. A., Fletcher, J. M., & Carter, S. (2015). Text-processing differences in adolescent adequate and poor comprehenders reading accessible and challenging narrative and informational text. *Reading Research Quarterly*, 50(4), 393-416. <https://doi.org/10.1002/rrq.105>
- Dürrenmatt, S. (2018). Ça chauffe ! Notre climat en sursis. *Coopération*, 1, 14-19.
- Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader-text interactions : how differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology*, 104(3), 515-528. <https://doi.org/10.1037/a0027182>
- Entin, E. B., & Klar, G. R. (1985). Relationships of measures of interest, prior knowledge, and readability to comprehension of expository passages. In B. A. Hutson (Ed.), *Advances in reading/language research* (pp. 9-38). JAI Press.



- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221-233. <https://doi.org/10.1037/h0057532>
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, 10(3), 323-330. https://doi.org/10.1207/s1532799xssr1003_7
- François, T. (2009, 24-26 juin). *Modèles statistiques pour l'estimation automatique de la difficulté de textes de FLE : Actes de la conférence TALN/RECITAL*. Senlis : ATALA.
- Giasson, J. (2016). *La compréhension en lecture* (3^e éd.). De Boeck Supérieur.
- Goigoux, R., & Thomazet, S. (1999). Pratiques et conceptions de l'enseignement de la lecture dans l'enseignement adapté. *Les Actes de lecture*, 67, 53-62.
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3), 371-395.
- Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.
- Habi, M. (2016). Le questionnaire en question. *Recherches*, 65, 109-127.
- Hiebert, E. H., & Pearson, P. D. (2010). *An examination of current text difficulty indices with early reading texts*. Textproject. http://textproject.org/assets/publications/TextProject_RRR-10.01_Text-Difficulty-Indices.pdf
- Joigneaux, C. (2009). La construction de l'inégalité scolaire dès l'école maternelle. *Revue française de pédagogie*, 169, 17-28. <https://doi.org/10.4000/rfp.1301>
- Kandel, L., & Moles, A. (1958). Application de l'indice de Flesch à la langue française. *Cahiers d'études de radio-télévision*, 19, 253-274.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension : a construction-integration model. *Psychological Review*, 95(2), 163-182.
- Kobayashi, K. (2009). The influence of topic knowledge, external strategy use, and college experience on students' comprehension of controversial texts. *Learning and Individual Differences*, 19(1), 130-134. <https://doi.org/10.1016/j.lindif.2008.06.001>
- Kobayashi, M. (2004). Reading comprehension assessment : from text perspectives. *Scientific Approaches to Language*, 3, 129-157.
- Lafontaine, D. (2003, 4-5 décembre). *Comment faciliter, développer et évaluer la compréhension des textes aux différentes étapes de la scolarité primaire ?* [Conference session]. Conférence de consensus sur l'enseignement de la lecture à l'école primaire, Paris, France.
- Laveault, D., & Grégoire, J. (2014). *Introduction aux théories de tests en psychologie et en sciences de l'éducation*. De Boeck Supérieur.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts : effects of prior knowledge and text coherence. *Discourse Processes*, 22(3), 247-288. <https://doi.org/10.1080/01638539609544975>
- McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Sources of text difficulty : across genres and grades. In J. P. Sabatini & E. Albro (Eds.), *Measuring up : advances in how we assess reading ability* (pp. 89-116). Rowman & Littlefield Education.
- Mesmer, H. A., Cunningham, J. W., & Hiebert, E. H. (2012). Toward a theoretical model of text complexity for the early grades : learning from the past, anticipating the future. *Reading Research Quarterly*, 47(3), 235-258. <https://doi.org/10.1002/rrq.019>
- Mesnager, J. (2002). Pour une étude de la difficulté des textes. *Le Français aujourd'hui*, 137(2), 29-40. <https://doi.org/10.3917/lfa.137.0029>
- Mesnager, J., & Bres, S. (2008). *Évaluer la difficulté des textes : cycle 2-3, 6e-5e*. [Computer software]. Nathan.
- Mosenthal, P. B., & Kirsch, I. S. (1998). A new measure for assessing document complexity : the PMose/Kirsch document readability formula. *Journal of Adolescent & Adult Literacy*, 41(8), 638-657.
- Mottier Lopez, L. (2008). *Apprentissage situé : la microculture de classe en mathématiques*. Peter Lang.
- Organisation for Economic Co-operation and Development (OECD). (2019). *PISA 2018 : assessment and analytical framework*. OECD Publishing. <https://doi.org/10.1787/b25efab8-en>
- Perrenoud, P. (2001). Evaluation formative et évaluation certificative : postures contradictoires ou complémentaires. *Formation professionnelle suisse*, 4, 25-28.
- Perrenoud, P. (2017). *Métier d'élève et sens du travail scolaire*. ESF sciences humaines.



- Pini, G., Reith, E., Weiss, L., & Bugniet, F. (2006). *Guide méthodologique pour l'évaluation et la mesure en éducation*. Développement et innovation pédagogique au cycle d'orientation (DIPCO).
- Rawson, K. A., & Kintsch, W. (2002). How does background information improve memory for text content ? *Memory & Cognition*, 30(5), 768-778. <https://doi.org/10.3758/bf03196432>
- Roth, M., De Pietro, J.-F., & Sánchez Abchi, V. (2014). Comment évaluer la lecture en référence au Plan d'études romand ? *Forumlecture.ch*, 3, 1-24.
- Roth, M., Sánchez Abchi, V., & De Pietro, J.-F. (2018). L'évaluation de la compréhension de l'écrit en français : comment estimer la difficulté des items et des textes proposés aux élèves ? In M. Demeuse, M. Milmeister & C. Weis (Eds.), *Actes du 30^{ème} colloque de l'ADMEE-Europe « L'évaluation en éducation et en formation face aux transformations des sociétés contemporaines », 10-12 janvier 2018, Campus Belval à Esch-sur-Alzette, Luxembourg* (pp. 468-469). Université du Luxembourg.
- Roth, M., Sánchez Abchi, V., & De Pietro, J.-F. (2019). Quel est l'impact de la difficulté des tâches évaluatives sur leur validité ? : symposium « Approches didactiques pour l'étude de la validité d'une évaluation : différents points de vue selon les disciplines ». In C. Gremion, E. Sylvestre & N. Younes (Dir.), *Actes du 31^{ème} colloque scientifique international de l'ADMEE-Europe « Entre normalisation, contrôle et développement formatif : évaluations sources de synergies ? », 8-10 janvier 2019, Université de Lausanne* (pp. 478-481). IFFP : CSE de l'Université de Lausanne.
- Sánchez Abchi, V., De Pietro, J.-F., & Roth, M. (2016). *Evaluer en français : comment prendre en compte la difficulté des items et des textes*. Institut de recherche et de documentation pédagogique.
- Weiss, J., & Wirthner, M. (1991). Evaluation de la compréhension de l'écrit correspondant au niveau de fin de deuxième année primaire. In J. Weiss & M. Wirthner (Eds.), *Enseignement du français : premiers regards sur une rénovation* (pp. 265-296). Delval.
- Zakharouch, J.-M. (2000). Les consignes au cœur de la classe : geste pédagogique et geste didactique. *Repères, recherches en didactique du français langue maternelle*, 22(1), 61-81. <https://doi.org/10.3406/reper.2000.2343>

Remerciements

Nous remercions Jean-François de Pietro qui a participé aux nombreuses réflexions et aux travaux qui ont permis la création de l'outil d'estimation de la difficulté et pour ses précieux conseils lors de l'élaboration de cet article. Nous remercions également Sylvie Küenzi pour les échanges autour de la question de la difficulté des textes et des questions et qui nous a fourni les documents analysés.

Murielle Roth

Collaboratrice scientifique à l'Institut de recherche et de documentation pédagogique (IRDP) à Neuchâtel. Chercheuse dans le domaine de l'évaluation des apprentissages et en didactique du français, ses travaux portent sur les pratiques d'évaluation en français langue première. Elle prépare une thèse de doctorat qui s'intéresse à la difficulté pour les élèves des textes et des questions exploités pour évaluer leur compréhension de l'écrit.

Verónica Sánchez
Abchi

Collaboratrice scientifique à l'IRDP à Neuchâtel et formatrice en didactique de l'espagnol à la Haute école pédagogique du Valais (HEPVS). Docteure en linguistique, ses travaux de recherche portent sur le domaine de la didactique et de l'évaluation des langues.

