

Appréciations et extraits des débats autour du texte soumis

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
Rédacteur-rices : Fernando MORALES VILLABONA &
Patricia SCHILLINGS

Ce fichier présente l'appréciation globale des deux rétroacteur-rices avant leurs échanges avec les autrices, puis quelques extraits des débats qui ont eu lieu autour du texte entre les autrices et les deux rétroacteur-rices. Ces extraits ont été choisis d'entente entre elles et eux.

Appréciations globales à propos du texte de la prépublication

Fernando MORALES VILLABONA

Le texte correspond au type de texte annoncé et est organisé selon les grandes parties qui le caractérisent :

Il s'agit bien d'un article de recherche empirique, structuré selon les grandes parties demandées.

- L'enchaînement des contenus du texte est cohérent (cohérence interne, fil rouge) : L'enchaînement des contenus est cohérent. La lecture est facilitée par les titres et sous-titres et le fil rouge du texte est aisément identifiable.

- Les questions et/ou hypothèses de recherche sont précisées : L'objet d'étude, en lien avec une problématique d'évaluation sommative, est présenté avec clarté. De même pour les questions de recherche, qui sont présentées déjà en introduction.

- Le texte propose un état argumenté des connaissances et des théories par rapport à l'objet d'étude :

Le cadre théorique présente bien un état des lieux des études portant sur l'objet du texte.

- La posture épistémologique est clairement indiquée : Une brève référence à la perspective située est faite dans l'avant-dernier paragraphe. S'agit-il bien de la perspective adoptée ? Une précision à ce propos serait la bienvenue.
- Les concepts-clés et/ou les variables étudiées sont explicitées et mises en relation : Les concepts-clés sont en effet mis en évidence et leur articulation est pertinente.
- Le contexte et les participant-e-s, la collecte de données et les outils associés sont précisés et appropriés : Des informations appropriées concernant la méthodologie de l'étude sont présentées. Or, une précision sur l'approche méthodologique servirait à guider davantage le lecteur. De plus, le statut des questionnaires récoltés dans quatre classes différentes mériterait d'être clarifié.
- La méthode d'analyse des données est décrite et elle est appropriée au questionnement de recherche : La méthode d'analyse annoncée est cohérente avec le questionnement de recherche.
- Les résultats sont interprétés avec rigueur et sont discutés au regard des objectifs de la recherche et de la littérature : Les résultats témoignent de la rigueur dans le travail d'analyse et de la cohérence avec le questionnement de recherche. La discussion des résultats revient sur les constats principaux, les élargit et les met en lien avec des éléments théoriques pertinents.
- Des perspectives de développement sont discutées et des retombées scientifiques et/ou sociales sont exposées : La conclusion de l'article clarifié le statut de l'étude, sa portée, et donne à voir des perspectives d'élargissement.
- Le texte respecte les normes éditoriales de La Revue LEeE : En grande partie, les normes éditoriales sont respectées. Or, des améliorations peuvent être apportées, par exemple, concernant l'utilisation de l'italique (vs les guillemets – ces derniers apparaissent sur-utilisés dans le texte), la numérotation des sous-titres (une occurrence) ou encore le langage épïcène (une occurrence). Les tableaux proposés sont lisibles, malgré une quantité importante d'informations, et offrent une vue synthétique des informations qui permet de mieux comprendre les propos.

Patricia Schillings

Le sujet de l'article est réellement intéressant, la revue de la littérature et la méthodologie sont clairement exposées. Même si j'ai éprouvé des difficultés à comprendre la partie résultats, je trouve certains de ceux-ci contre intuitifs et donc susceptibles de faire avancer la réflexion en matière d'enseignement de la lecture.

Les critères ci-dessus me paraissent atteints : néanmoins, la formulation des questions de recherche évolue au fil du texte et la question des conceptions de l'enseignante bien développée dans les résultats n'apparaît pas en tant que telle en fin d'introduction. Les variables étudiées sont explicitées mais la flexibilité cognitive du lecteur est mise à rude épreuve dans la présentation des résultats qui pourrait être plus précise et constante au niveau des modes de présentation des données.

Extraits des débats autour du texte soumis

...

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 âgé-es de 12-13 ans) mise en oeuvre 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, De Pietro et Roth (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, De Pietro and Roth (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 allons nous intéresser à une pratique d'évaluation sommative de la compréhension de l'écrit mise en oeuvre 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

Commenté [1]: PATRICIA SCHILLINGS : Cette étude serait me semble-t-il à mettre en lien avec l'approche psycho-didactique de la validité défendue par Vantourout & Goasdoué (2014). en effet ce courant consiste notamment à analyser les démarches cognitives mises en jeu dans les épreuves.

Commenté [2]: MURIELLE ROTH ET VERONICA SÁNCHEZ ABCHI : Nous connaissons les travaux de cette équipe et nous avons plusieurs points communs. La différence centrale est que, même s'il y a un intérêt pour le contenu évalué et la tâche, le but est différent. Ils s'intéressent à la validité des évaluations. De notre côté, nous cherchons à mettre à l'épreuve un outil pour estimer la difficulté des textes et des questions en CE. Nous touchons à des questions de validité mais sans que cela soit l'objectif de départ.

Commenté [3]: PATRICIA SCHILLINGS : Je comprends bien que votre focus porte sur la mesure de la difficulté.

Commenté [4]: PATRICIA SCHILLINGS : Cette conduite d'ajustement au niveau des élèves est au cœur de l'analyse sociologique de l'échec scolaire menée par Joignaux et Bonnéry (2009). L'ajustement contribuerait à accentuer les inégalités de départ lorsqu'il propose un cadrage trop serré des tâches aux élèves plus faibles (découpage des tâches, questions plus faciles), ce qui les dispenserait d'apprendre de la complexité des tâches. Je me demande dans quelle mesure vos résultats ne pourraient pas rejoindre ce courant de recherche.

Commenté [5]: MURIELLE ROTH ET VERONICA SÁNCHEZ ABCHI : Nous n'avons pas fait référence à cette étude, car elle relève d'un autre cadre théorique que nous n'avons pas pris en compte.

Commenté [6]: PATRICIA SCHILLINGS : Je trouve néanmoins dommage dans la discussion de ne pas ouvrir la réflexion sur les causes et les effets potentiels de l'ajustement du niveau de difficulté des épreuves aux niveaux des élèves ...

Commenté [7]: MURIELLE ROTH ET VERONICA SÁNCHEZ ABCHI : Vous avez raison d'insister sur ce point. Nous avons pris le temps d'en parler et bien que nous n'ayons pas été sur le terrain des inégalités scolaires, nous pouvons tout à fait ouvrir la réflexion dans la discussion. Ainsi, vous trouverez quelques lignes à ce sujet dans le chapitre concerné.



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, De Pietro et Roth (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; Perreroud, 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 Landsheer, 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, De Pietro et Roth (2016) ont mené une étude exploratoire qui les a conduits 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.

Commenté [8]: PATRICIA SCHILLINGS : En fait vous cherchez à identifier quelles variables sont fonctionnelles aux yeux de l'enseignante, cette manière d'entrer dans le système interprétatif, si elle était complétée par des entretiens avec la professionnelle vous rapproche de la didactique professionnelle.

Commenté [9]: MURIELLE ROTH ET VERONICA SÁNCHEZ ABCHI : Nous n'avons malheureusement pas pu réaliser des entretiens avec l'enseignante qui a créé les questionnaires, c'est pourquoi nous ne pouvons pas nous situer dans le cadre théorique de la didactique professionnelle même si des parallèles peuvent être faits.

Commenté [10]: PATRICIA SCHILLINGS : bien d'accord, c'était juste une réflexion suscitée par vos analyses qui lèvent quand même un coin du voile sur l'activité d'évaluation de l'enseignante (que vous nommez pratique d'évaluation).



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 en cohérence avec l'outil de Sánchez Abchi, De Pietro et Roth (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 enseignants. Plus spécifiquement, nous souhaitons répondre aux 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 différentes ?

2. Le cadre de référence

aidera à la compréhension si la question implique la mobilisation d'un vocabulaire.

2. Méthodologie

2.1 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. Nous nous sommes intéressées en particulier à ses pratiques d'évaluation, car elle a pour habitude de se questionner sur le contenu des épreuves qu'elle élabore pour ses élèves. De plus, 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.

2.2 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 *Ca chauffe ! Notre climat en sursis* (Dürrenmatt, 2018). Il appartient au regroupement de genre *qui transmet des savoirs* (Conférence intercantonale

Commenté [11]: FERNANDO MORALES VILLABONA : C'est très intéressant. Je me demande toutefois : qu'est-ce qui justifie cette comparaison ? Est-ce à des fins de validation de l'outil ? ou pour se prononcer sur les liens (ou les ruptures) entre l'outil (construit sur la base de l'analyse d'épreuves externes) et les pratiques d'évaluation interne ?

Commenté [12]: MURIELLE ROTH ET VERONICA SANCHEZ ABCHI : Nous avons apporté des précisions sur l'enjeu de cette comparaison.

Commenté [13]: FERNANDO MORALES VILLABONA : Merci pour cette précision, que je trouve importante

Commenté [14]: PATRICIA SCHILLINGS : Je me demande si l'analyse des conceptions de l'enseignante est une question de recherche additionnelle ou si elle découle de la première question ou de l'analyse des ajustements.

Commenté [15]: MURIELLE ROTH ET VERONICA SANCHEZ ABCHI : Nous pensons que c'est l'analyse des ajustements qui vont nous donner des indices sur la manière dont l'enseignante envisage l'enseignement-apprentissage. Il ne s'agit donc pas d'une question additionnelle.

Commenté [16]: PATRICIA SCHILLINGS : ok

Commenté [17]: FERNANDO MORALES VILLABONA : Ce paragraphe me semble un peu déconnecté. Par exemple, la référence à des documents analysés, sans avoir d'autres informations sur la démarche d'analyse prévue/effectuée est difficile à comprendre. Une ou deux phrases introductives pourraient donner des éléments de contexte pour le lecteur.

Une deuxième difficulté est associée à la référence aux copies recueillies dans quatre classes différentes, alors qu'avant (§ 3.1) il est annoncé que le regard portera sur les pratiques d'une seule enseignante. Une clarification semble donc nécessaire à ce propos.

Commenté [18]: MURIELLE ROTH ET VERONICA SANCHEZ ABCHI : Merci pour le commentaire. Nous avons réécrit et déplacé un paragraphe pour plus de clarté concernant le corpus.

Commenté [19]: FERNANDO MORALES VILLABONA : C'est bien plus clair avec cette organisation par sous-titres, merci

Commenté [20]: PATRICIA SCHILLINGS : plus clair en effet !



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, le B 18 questions, et le C 14 questions. 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.

2.3 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 différent-elles? –, 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.

2.4 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 <i>Texte</i>			Indice <i>Contenu questions</i>			Indice <i>Enveloppe questions</i>		
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 <i>Texte</i> convertie en indice qualitatif en se référant au Tableau 1 puis pondération par bonus et malus			Valeur moyenne <i>Contenu questions</i> convertie en indice qualitatif en se référant au Tableau 1			Valeur moyenne <i>Enveloppe questions</i> convertie en indice qualitatif en se référant au Tableau 1		
Indice qualitatif <i>Texte</i>			Indice qualitatif <i>Contenu questions</i>			Indice qualitatif <i>Enveloppe questions</i>		

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
---------	---------------------------------

Commenté [21]: PATRICIA SCHILLINGS : Cette partie m'a demandé plusieurs relectures successives car j'ai éprouvé des difficultés à saisir les liens entre les deux tableaux. Je me trompe peut-être mais pour comprendre vos résultats, il est plus important de comprendre l'idée que les indices reposent sur une moyenne de trois dimensions (tab 2) que la manière, assez complexe dont vous passez d'une échelle ordinaire en trois niveaux à une moyenne associée à une échelle ordinaire en 5 niveaux. Ces aspects techniques me paraissent secondaires par rapport au fait que vous appréhendez la difficulté de manière qualitative (en décrivant les dimensions) et quantitative en y associant des indices chiffrés. J'inverserais donc les deux tableaux ...

Commenté [22]: MURIELLE ROTH ET VERONICA SÁNCHEZ ABCHI : Merci pour votre commentaire. Nous avons inversé les deux tableaux et tenté de clarifier les choses.

Commenté [23]: PATRICIA SCHILLINGS : Cela me semble à présent plus cohérent pour un lecteur externe. Merci



-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	

3. Résultats

''''

4. 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, De Pietro et Roth (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.

Commenté [24]: PATRICIA SCHILLINGS : Ce résultat contre-intuitif est sans doute le plus intéressant de votre étude ! Personnellement je m'attendais à ce que le questionnaire C comporte une proportion nettement plus importante de questions spécifiques ciblées sur des processus de bas niveau et à l'inverse une proportion nettement plus importante de questions ciblées sur l'interprétation dans le questionnaire A . Cette logique inconsciente ? de simplification de la tâche a été mise en évidence dans une étude de Goigoux et Thomazet menée en France . Cette analyse du travail des enseignants interrogés montre une interaction entre des pratiques pédagogiques en matière de lecture et les caractéristiques des élèves. Leur travaux montrent que les conceptions erronées développées par les élèves à propos de la lecture émanait des pratiques enseignantes.

Goigoux, Serge Thomazet. Pratiques et conceptions de l'enseignement de la lecture dans l'enseignement adapté. Les Actes de Lecture, 1999, 67, pp.53-62.

Commenté [25]: MURIELLE ROTH ET VERONICA SÁNCHEZ ABCHI : Merci de nous avoir signalé cette référence. Nous nous y sommes référés.

