

Évaluer l'oral quand on est enseignant ou chercheur : points de discussion et prises de décision dans la coconception d'une grille critériée

Christine Wiertz, Sabine Van Mosnenck, Benoit Galand and Stéphane Colognesi

Volume 43, Number 3, 2020

URI: <https://id.erudit.org/iderudit/1083006ar>

DOI: <https://doi.org/10.7202/1083006ar>

[See table of contents](#)

Publisher(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (print)

2368-2000 (digital)

[Explore this journal](#)

Cite this article

Wiertz, C., Van Mosnenck, S., Galand, B. & Colognesi, S. (2020). Évaluer l'oral quand on est enseignant ou chercheur : points de discussion et prises de décision dans la coconception d'une grille critériée. *Mesure et évaluation en éducation*, 43(3), 1–37. <https://doi.org/10.7202/1083006ar>

Article abstract

Assessment rubrics seem to be a relevant solution when teachers and researchers have difficulties in assessing oral language effectively. However, the interpretation of the rubrics depends on the evaluator's frame of reference. In this study, a comparison was made between a group of five teachers and a group of four researchers in the conception of two rubrics to assess pupils' oral communication skills at the end of elementary school. The results of our qualitative analysis show that the difficulties, the ways of using the rubrics, the types of justifications and the solutions experimented by the two groups of professionals somewhat vary. Teachers opted for more grounded solutions, strongly linked to their experience and their classroom context. As for researchers, they had to make sure that the constructed rubric allows a precise measure of oral communication skills and had to question the structural organization of the rubric.

Évaluer l'oral quand on est enseignant ou chercheur : points de discussion et prises de décision dans la coconception d'une grille critériée

Christine Wiertz

Université catholique de Louvain

Sabine Van Mosnenck

Haute École Galilée

Benoit Galand

Stéphane Colognesi

Université catholique de Louvain

MOTS CLÉS: évaluation de l'oral, grille critériée, processus de coconception, recherche collaborative

Devant les difficultés des enseignants et des chercheurs à évaluer efficacement l'oral, les grilles critériées semblent une solution pertinente. Cependant, l'interprétation des critères de la grille varie en fonction du cadre de référence de l'évaluateur. Dans cette étude, une comparaison a été réalisée entre un groupe de cinq enseignants et un groupe de quatre chercheurs dans la conception de deux grilles critériées permettant d'évaluer la compétence à communiquer oralement des élèves de fin de primaire. Les résultats de notre analyse qualitative montrent que les difficultés, les modalités d'utilisation de la grille, les types de justifications et les solutions apportées par les deux groupes professionnels varient quelque peu. Les propos des enseignants sont plus ancrés dans le concret et fortement liés au contexte de la classe. Le groupe des chercheurs, quant à lui, s'est assuré que l'outil construit permet une mesure fine de la compétence à communiquer oralement et s'est interrogé sur l'organisation structurelle de la grille construite.

KEY WORDS: oral evaluation, evaluation rubric, co-conception process, collaborative research

Assessment rubrics seem to be a relevant solution when teachers and researchers have difficulties in assessing oral language effectively. However, the interpretation of the rubrics depends on the evaluator's frame of reference. In this study, a comparison was made between a group of five teachers and a group of four researchers in the conception of two rubrics to assess pupils' oral communication skills at the end of elementary school. The results of our qualitative analysis show that the difficulties, the ways of using the rubrics, the types of justifications and the solutions experimented by the two groups of professionals somewhat vary. Teachers opted for more grounded solutions, strongly linked to their experience and their classroom context. As for researchers, they had to make sure that the constructed rubric allows a precise measure of oral communication skills and had to question the structural organization of the rubric.

PALAVRAS-CHAVE: avaliação oral, rubrica, processo de coconceção, investigação colaborativa

Perante as dificuldades de professores e investigadores em avaliar eficazmente a oralidade, as rubricas parecem ser uma solução pertinente. No entanto, a interpretação dos critérios de uma rubrica varia em função do quadro de referência do avaliador. Neste estudo, foi feita uma comparação entre um grupo de cinco professores e um grupo de quatro investigadores na conceção de duas rubricas para avaliar a competência de comunicação oral de alunos no final do 1.º ciclo do ensino básico. Os resultados da nossa análise qualitativa mostram que as dificuldades, as modalidades de utilização da rubrica, os tipos de justificações e as soluções apresentadas pelos dois grupos profissionais variam um pouco. Os comentários dos professores estão mais ancorados no concreto e fortemente vinculados ao contexto da sala de aula. O grupo de investigadores, por sua vez, certificou-se de que a ferramenta construída permite uma medição fina da competência para se comunicar oralmente e questionou-se sobre a organização estrutural da rubrica construída.

Note des auteurs : La correspondance liée à cet article peut être envoyée à christine.wiertz@uclouvain.be. Nous remercions le Groupe collaboratif pour l'enseignement de l'oral (GCEO) pour son investissement dans cette recherche.

Problématique

L'oral occupe une place majeure aussi bien dans la vie quotidienne des élèves (Lafontaine, 2016) que dans l'enseignement, qui se fait en grande partie par le biais du canal de l'oral (Allen, 2017 ; Schneuwly, 1997). La non-maîtrise de l'oral peut donc être à l'origine d'une surcharge cognitive qui compliquera la compréhension de l'élève des enseignements donnés et ainsi son entrée dans le processus d'apprentissage (Allen, 2017 ; Sweller et al., 2011).

Malgré la place importante prise par l'oral dans la vie des élèves, celui-ci reste actuellement le parent pauvre de la didactique du français (Gagnon et al., 2017 ; Nonnon, 2016). Par exemple, très peu de recherches s'attardent à la progression développementale typique des élèves en *savoir parler* et en *savoir écouter*¹ au primaire (Garcia-Debanc, 1999 ; Nonnon, 2016). Ce vide théorique complique la création de séquences didactiques et de critères d'évaluation adaptés à l'âge des élèves (Colognesi et Dolz-Mestre, 2017 ; Nonnon, 2016), tant pour les enseignants que pour les chercheurs.

Plus largement, les informations relatives aux différentes spécificités constituant l'oral sont difficiles d'accès pour les acteurs du terrain (Dumais, 2014 ; Dumais et Lafontaine, 2011) et peu partagées dans la formation initiale des enseignants (Colognesi et Deschepper, 2019). De fait, il n'est pas simple pour les (futurs) enseignants de savoir ce qu'ils peuvent/doivent enseigner et donc évaluer lorsqu'ils envisagent de travailler la compétence à communiquer oralement² de leurs élèves.

D'autant plus que l'oral est indissociable de la personne que représente l'orateur (Alrabadi, 2011 ; Garcia-Debanc, 1999), de son corps, de sa voix (Berdal-Masuy et Renard, 2015 ; Garcia-Debanc, 1999) et de son identité (Garcia-Debanc, 1999 ; Maurer, 2001). Couplé à l'aspect volatile de l'oral et au manque de connaissance concernant ces caractéristiques, tout ceci peut rendre son évaluation subjective (Garcia-Debanc, 1999). De plus, l'enjeu est de ne pas évaluer l'oral simplement comme un tout, mais de bien différencier ses composantes propres (Lafontaine et Préfontaine, 2007).

Une voie à explorer pour faire face à ces obstacles et pour faciliter l'évaluation de la prise de parole serait les grilles critériées (Dunbar et al., 2006 ; Lafontaine et Messier, 2009). Cependant, celles-ci sont encore peu nombreuses à être mises à disposition des évaluateurs (Alrabadi, 2011 ; Lafontaine et Messier, 2009) et aussi peu documentées dans la littérature. De plus, pour que leur utilisation soit efficace, il est nécessaire que les critères soient bien compris et appréhendés de manière similaire par l'ensemble des évaluateurs, ce qui n'est pas toujours le cas (Balan et Jönsson, 2018 ; Baume et al., 2004).

L'objectif de cette étude est de contribuer à répondre à cette difficulté d'évaluer la compétence à communiquer oralement des élèves. Pour ce faire, cette étude exploratoire, ancrée dans une recherche collaborative, vise à analyser les discussions et thèmes abordés par des groupes d'enseignants et de chercheurs au sein d'un processus de coconception de grilles d'évaluation. Plus précisément, nous tentons de répondre à deux questions : (a) Quels sont les points de discussion mis en avant par les enseignants et par les chercheurs lors d'un processus de coconception d'une grille critériée évaluant l'oral ? (b) Quelles décisions sont prises pour ajuster la grille ?

Cadre conceptuel

Nous abordons d'abord les différentes modalités que peuvent recouvrir les composantes de la compétence à communiquer oralement. Ensuite, les avantages et les difficultés relatifs à l'utilisation des grilles critériées sont identifiés. Enfin, nous terminons par mettre en évidence en quoi la discussion au sein d'un groupe permet de créer des outils d'évaluation plus adaptés à la réalité professionnelle des différents membres.

Statut et composantes de l'oral

En contexte scolaire, l'oral peut intervenir dans différents types de situations scolaires, qui sont au moins de quatre ordres (Plane, 2015). Premièrement, l'oral intervient dans les situations ordinaires de communication. Deuxièmement, il a lieu dans les situations pédagogiques où sont mobilisées les interactions de la vie quotidienne, par exemple la transmission et l'écoute d'informations, de consignes ou encore lors d'échanges de type questions/réponses (Colognesi et Dolz-Mestre, 2017 ; Lafontaine et Messier, 2009). Considéré comme un *médium* (Dumais et Lafontaine, 2011), il désigne alors une modalité pédagogique (Plane, 2015).

C'est actuellement cette modalité qui est la plus souvent retrouvée au sein des classes (Colognesi et Deschepper, 2019 ; Lafontaine et Messier, 2009). Troisièmement, l'oral peut être considéré comme un *outil d'enseignement et d'apprentissage*, dans les disciplines. On pense alors au rôle de la verbalisation et des interactions dans la construction des apprentissages, soit l'oral réflexif (Allen, 2017 ; Chabanne et Bucheton, 2002 ; Vygotski, 1934/2013) dans une perspective métacognitive (Berger et Büchel, 2013). Quatrièmement, l'oral peut être lui-même un *objet d'apprentissage*. C'est ici que se situe le sujet de notre étude.

Cela sous-entend la mise en place, dans la grille horaire, de séquences didactiques (Colognesi et Dolz-Mestre, 2017 ; Dolz et Schneuwly, 1998 ; Dupont, 2019) dédiées au développement de la compétence à communiquer oralement et à son enseignement spécifique (Dupont et Grandaty, 2016). Par ailleurs, ces diverses situations scolaires ne sont pas toujours indépendantes les unes des autres et un certain recoupement peut avoir lieu, ce qui complexifie de ce fait la mise en exergue d'une conceptualisation claire de l'oral (Plane, 2015) et de son évaluation (Garcia-Debanco et Delcambre, 2001).

Une amélioration dans les différentes composantes de l'oral est observée chez les élèves lorsque l'oral est abordé comme un objet d'enseignement (Colognesi et al., 2020 ; Dolz et Gagnon, 2008 ; Dolz et Schneuwly, 1998 ; Dumais et al., 2015). Dans ce cas de figure, trois capacités langagières sont à considérer selon Dolz et al. (1993) : (a) la *capacité d'action* : capacité d'adapter son discours à la situation de communication en respectant les codes structurels d'un type de discours ; (b) la *capacité discursive* : habiletés du locuteur à produire du contenu, tout en l'organisant de manière structurée, claire et cohérente et (c) la *capacité linguistico-discursive* : l'utilisation d'un vocabulaire correct et adapté à la situation de communication ainsi que la mobilisation des règles grammaticales et syntaxiques appropriées.

Dans notre contexte, les *Socles de compétences – français* (Fédération Wallonie-Bruxelles, 2013) mettent en évidence quatre compétences orales. Bien que, conceptuellement, la capacité, plus décontextualisée, peut se différencier de la compétence, associée à une situation complexe et précise (Perrenoud, 2000), un parallèle relatif peut néanmoins être mis en place (Colognesi et Hanin, 2020).

La première compétence est le fait d'*orienter sa parole et son écoute en fonction de la situation de communication*. Il s'agit de pouvoir identifier les spécificités du discours telles que les caractéristiques du destinataire, le contexte communicationnel, l'objectif du locuteur, etc. Nous pouvons relier cette compétence à la capacité d'action de Dolz et al. (1993).

La deuxième compétence, *élaborer des significations*, regroupe les éléments relatifs au contenu du discours qui peuvent être associés à la capacité discursive.

La troisième compétence, *assurer et dégager l'organisation et la cohérence du message (en utilisant le lexique et la morphosyntaxe)*, met en exergue l'importance d'organiser son discours et d'en faire un tout cohérent, tout en respectant une syntaxe et un lexique appropriés. Cette compétence peut également être rattachée à la capacité discursive, mais aussi à la *capacité linguistico-discursive*.

La dernière compétence répertoriée dans le programme belge peut également, en partie, être reliée à la capacité linguistico-discursive. Il s'agit de la compétence *utiliser et identifier les moyens non verbaux*, qui peut être explicitée par le fait d'arriver à soutenir de manière cohérente son discours au moyen de supports physiques, mais également par l'utilisation de la voix et du corps.

Tableau 1
*Parallèle entre les capacités langagières de Dolz et al. (1993)
 et les compétences des Socles de compétences – français (FW-B, 2013),
 repris de Colognesi et Hanin (2020)*

Capacités langagières (Dolz et al., 1993)	Compétences orales des Socles de compétences (FW-B, 2013)
Capacité d'action	<i>C1. Orienter sa parole et son écoute en fonction de la situation de communication</i>
Capacité discursive	<i>C2. Élaborer des significations</i> <i>C3. Assurer et dégager l'organisation et la cohérence du message (en utilisant le lexique et la morphosyntaxe)</i>
Capacité linguistico-discursive	<i>C3. Assurer et dégager l'organisation et la cohérence du message (en utilisant le lexique et la morphosyntaxe)</i> <i>C4. Utiliser et identifier les moyens non verbaux</i>

Le Tableau 1 propose une mise en relation des capacités langagières avec les compétences mobilisées. Un défi majeur pour les enseignants et les chercheurs est cependant de savoir comment évaluer ces différentes composantes de l'oral.

Grilles critériées : un moyen pour évaluer l'oral ?

Malgré différents travaux ayant pour objectifs de clarifier ce qu'est l'oral (Dumais, 2016), la difficulté des enseignants à se représenter clairement le *savoir parler* et ses composantes représente une contrainte majeure pour évaluer finement l'oral (Colognesi et Deschepper, 2019 ; Colognesi et Dolz-Mestre, 2017 ; Dumais et Lafontaine, 2011). Les grilles critériées seraient une des solutions possibles pour pallier cela (Dunbar et al., 2006), dans une optique tant formative que certificative (Berthiaume et al., 2011).

Comme le mettent en évidence Berthiaume et al. (2011), une grille critériée peut être conceptualisée comme étant un tableau reprenant les différentes dimensions qui seront évaluées par l'enseignant. Chacune de ces dimensions comprend différents critères. Pour chaque critère, une échelle d'appréciation permet à l'évaluateur d'établir le niveau de performance de l'élève.

Les grilles critériées ont l'avantage d'offrir aux évaluateurs un système d'évaluation analytique dans lequel chaque composante est évaluée individuellement et de manière similaire pour chaque individu (Berthiaume et al., 2011). Elles permettent d'obtenir un jugement plus valide, équitable et fiable de la performance de l'élève (Dunbar et al., 2006 ; Pomplun et al., 1998). Par ailleurs, l'établissement des critères finaux peut soutenir les enseignants à se représenter concrètement les objectifs d'apprentissage, ce qui peut les aider à structurer et à préparer leurs cours (Berthiaume et al., 2011) et à assurer l'alignement pédagogique (Biggs, 1996). Les grilles critériées peuvent également servir de base au feedback donné aux élèves (Balan et Jönsson, 2018 ; Berdal-Masuy et Renard, 2015) et leur permettre de mieux visualiser l'écart entre leur performance et ce qui est attendu par l'évaluateur (Berdal-Masuy et Renard, 2015 ; Berthiaume et al., 2011).

Cependant, il n'existe que très peu de grilles critériées permettant d'évaluer les différentes composantes de la compétence à communiquer oralement (Alrabadi, 2011 ; Lafontaine et Messier, 2009) telles qu'elles ont été présentées *supra*. Effectivement, bien que des grilles permettant d'évaluer la capacité linguistico-discursive ainsi que les comportements verbaux

et paraverbaux soient disponibles (Berdal-Masuy et Renard, 2015; Dunbar et al., 2006; Gregg et Mather, 2002), très peu d'outils se concentrent également sur la capacité d'action et sur la capacité discursive.

Par ailleurs, la plupart des critères utilisés dans les grilles s'inspirent de ce qui peut être attendu d'une prestation écrite et, partant, ne permettent pratiquement pas de prendre en compte les spécificités de l'oral (Alrabadi, 2011). Plus largement, les différents prescrits ministériels dans notre contexte ne proposent, à notre connaissance, aucune grille critériée permettant d'évaluer les compétences orales reprises dans le programme. Ce faisant, il est fréquent que les enseignants créent leur propre outil d'évaluation en s'appuyant sur les compétences reprises dans les programmes (Lafontaine et Messier, 2009).

Bien que les grilles critériées permettent d'augmenter la fiabilité de l'évaluateur dans son appréciation de la prestation de l'élève (Berthiaume et al., 2011), elles ne permettent pas de supprimer toute part de subjectivité (Berthiaume et al., 2011; Dunbar et al., 2006). En effet, chaque évaluateur est influencé par son propre cadre de référence, construit inconsciemment au fil de ses expériences personnelles (Coen et al., 2008; Coen et Pellegrini, 2011). Les utilisateurs de la grille n'ont également pas toujours le niveau d'expertise suffisant pour comprendre et employer l'outil comme prévu par ses créateurs (Balan et Jönsson, 2018). Ces différents constats expliquent en partie le fait que le coefficient de fidélité interjuges entre deux utilisateurs d'une grille critériée évaluant l'oral tend à ne pas être très élevé (Dunbar et al., 2006; Pitts et al., 2002).

Par ailleurs, il semble assez complexe de traduire une compétence en critères opérationnalisables (Bouwer et al., s. d.) et d'établir un score seuil permettant de déterminer objectivement si l'élève atteint ou non chacun des critères de la grille (Derycke, 2000). De plus, la somme des différents critères ne permet pas toujours de refléter efficacement la prestation globale de l'élève ni d'obtenir une évaluation complète de ce qui est mesuré (Bouwer et al., s. d.; Derycke, 2000).

La diminution des échelons de l'échelle d'évaluation ainsi que l'affinement de leur signification sont des premières pistes de solution permettant d'augmenter la similarité de réponse entre différents évaluateurs (Dunbar et al., 2006). Une deuxième possibilité est de permettre aux évaluateurs de discuter entre eux de la signification des différents critères afin qu'ils puissent se créer des standards similaires, ce qui augmente ainsi la fidélité

interjuges (Dunbar et al., 2006; Pitts et al., 2002). Cette solution permet également une compréhension de la grille plus uniforme et une meilleure appropriation des concepts sous-tendant la grille (Berthiaume et al., 2011; Dunbar et al., 2006).

Notons toutefois que d'autres outils d'évaluation existent au-delà des grilles, par exemple le dossier d'apprentissage (Derycke, 2000), les méthodes holistiques ou encore les productions de référence (Bouwer et al., s. d.). Ils permettent de pallier certaines difficultés relatives à l'utilisation d'une grille critériée en augmentant notamment la validité de l'évaluation et sa rapidité d'exécution (Bouwer et al., s. d.). Cependant, comme nous l'avons expliqué *supra*, pour un objet comme l'oral dont les différentes composantes restent floues pour la plupart des évaluateurs, l'utilisation d'une grille critériée permet d'indiquer clairement les différents points d'attention à avoir. C'est pour cette raison que nous avons choisi de nous concentrer sur cet outil, tout en restant attentifs à ses désavantages et options de rechange.

Discussions entre évaluateurs : un moyen pour améliorer les grilles critériées ?

La discussion est un processus dynamique qui peut provoquer des négociations de sens entre les différents participants (Gan et al., 2009). C'est le cas notamment au sein d'un conflit sociocognitif, où un groupe de pairs d'avis contraire est amené à discuter afin de résoudre un problème (Buchs et al., 2008). Dans cette optique, il est nécessaire qu'il y ait des discussions, des malentendus ainsi que des désaccords entre les participants pour que les représentations de chacun se confrontent à celles des autres (Vinatier, 2010). Ainsi, les acteurs de la conversation sont poussés à argumenter et à expliciter leurs points de vue, ce qui peut stimuler la production d'idées créatives et de qualité (Morrissette et al., 2012). La prise en compte des différents points de vue permet également d'augmenter le degré d'expertise de chacun par rapport au thème abordé (Buchs et al., 2008; Morrissette et al., 2012).

Cependant, ces conséquences positives ne se réalisent pas systématiquement (Buchs et al., 2008; Darnon et al., 2006). En effet, elles sont plus marquées lorsque les participants se concentrent sur la tâche en suivant une régulation épistémique qui leur permet d'intégrer l'opinion de chacun, plutôt qu'une régulation relationnelle centrée sur la comparaison et la sélection d'une seule bonne réponse (Buchs et al., 2008; Darnon et al., 2006).

Plus largement, dans une démarche de révision d'un outil, les discussions entre acteurs de terrain font émerger de multiples savoirs d'expérience (Mottier Lopez, 2015b) qui participent à un meilleur ajustement de l'outil à la réalité professionnelle des membres de la discussion (Morrissette et al., 2012). Ces savoirs d'expérience sont spécifiques à chacun (Coen et Pellegrini, 2011 ; Morrissette et al., 2012). Cependant, certaines similitudes peuvent être retrouvées en fonction du groupe professionnel (Jorro, 2009). Par exemple, les chercheurs tendent à apporter des éléments de réflexion autour des cadres théoriques et des procédures méthodologiques (Morrissette et al., 2012). Les enseignants vont plutôt se référer aux contraintes concrètes du système scolaire et aux expériences vécues en interaction avec les élèves et leurs collègues (Morrissette et al., 2012). Cette intégration des différents cadres de référence et savoirs expérientiels constitue notamment une caractéristique importante des recherches collaboratives (Desgagné et Larouche, 2010 ; Mottier Lopez, 2015a, 2015b ; Vinatier, 2010).

Pour que ces partages puissent avoir lieu, en plus du contrat explicite qui lie les partenaires sur leurs modes de fonctionnement et des buts à atteindre, deux aspects sont nécessaires. Premièrement, le *contrat réflexif*, mis en évidence par Bednarz et al. (2012). D'après les auteurs, il s'agit de :

règles plus implicites [qui] se construisent dans l'interaction, au fil des rencontres, en fonction des partenaires, de leurs attentes, des rapports qu'ils établissent entre eux, en fonction, somme toute, du besoin de se comprendre, de coordonner ses efforts, d'une rencontre à l'autre, vers une cible qui elle-même comporte une part d'indétermination et qui se négocie au fil des interactions des partenaires (p. 6).

Deuxièmement, pour justement atteindre l'objectif, un dispositif de *modération sociale* (Mottier Lopez et al., 2012) permet le consensus nécessaire à l'intégration des différents apports.

Objectifs de cette étude

Comme nous l'avons expliqué *supra*, les grilles critériées sont un moyen parmi d'autres d'évaluer les différentes composantes de l'oral, avec l'avantage que les critères, s'ils sont bien construits et formulés, permettent de clarifier les aspects auxquels l'évaluateur doit être attentif (Dunbar et al., 2006). Cependant, l'interprétation et l'utilisation des critères peuvent varier d'un évaluateur à l'autre (Dunbar et al., 2006) puisque chacun est influencé par son cadre de référence, par son terrain professionnel (Coen et

al., 2008 ; Coen et Pellegrini, 2011 ; Jorro, 2009) et par son niveau d'expertise (Balan et Jönsson, 2018). La discussion (Dunbar et al., 2006 ; Pitts et al., 2002) semble une piste intéressante pour pallier ces difficultés et pour adapter la grille aux contraintes professionnelles de chacun.

Cette étude, à visée compréhensive, est ancrée dans les sciences de l'éducation et dans la didactique du français. D'une part, son objectif est d'en apprendre plus sur les points d'attention d'un groupe d'enseignants et de chercheurs dans un processus de coconception d'une grille critériée permettant d'évaluer les composantes de la compétence à communiquer oralement. D'autre part, elle vise à mieux comprendre les décisions qui sont prises au sein des groupes pour améliorer la grille. La visée à plus long terme de ce projet est de réaliser une cartographie des compétences orales des élèves en Fédération Wallonie-Bruxelles afin d'identifier les difficultés des élèves dans ce domaine et de les soutenir plus efficacement.

Pour y arriver, la création d'outils d'évaluation de l'oral utiles, utilisables et acceptables (Tricot et al., 2003) aussi bien dans le domaine de l'enseignement que de la recherche semble une étape indispensable. Analyser les points de tension et d'accord d'un groupe d'enseignants et d'un groupe de chercheurs quant à l'utilisation d'une grille critériée est la première étape que nous suivons pour mieux comprendre les obstacles rencontrés lorsqu'ils utilisent cet outil pour évaluer une performance orale. Cette compréhension permettrait de concevoir des outils plus en adéquation avec les besoins de ces deux corps de métiers.

Méthodologie

Contexte de l'étude

Cette étude s'inscrit dans un projet plus large au sein duquel plusieurs étapes (explicitées ci-après) ont été suivies pour construire des grilles critériées permettant d'évaluer la compétence à communiquer oralement. Il s'agit d'une recherche collaborative (Desgagné et Larouche, 2010 ; Van Nieuwenhoven et Colognesi, 2015) au sein du Groupe collaboratif pour l'enseignement de l'oral (GCEO). Comme tout groupe collaboratif, le GCEO a pour objectif de faire émerger une activité de production de connaissances, mais aussi de permettre le développement professionnel de chaque catégorie d'acteurs (Bourassa et al., 2007). Ce groupe rassemble 15 participants : 6 formateurs d'enseignants (didacticiens

et psychopédagogues), 5 enseignants et 4 chercheurs universitaires de la Fédération Wallonie-Bruxelles de Belgique (3 spécialisés en sciences de l'éducation et 1 en didactique du français). Bien que tous les membres du groupe aient été sollicités pour un travail sur les grilles critériées, nous nous concentrons ici uniquement sur les données relatives aux enseignants et aux chercheurs, qui sont pertinentes pour nos questions de recherche.

Participants

Deux sous-groupes ont fonctionné en parallèle dans un processus de coconception d'une grille d'évaluation, explicité dans la section suivante. Le groupe d'enseignants est composé de quatre institutrices au primaire et d'une enseignante de français au secondaire inférieur³. Le groupe de chercheurs est constitué de trois femmes (deux professeures et une doctorante) et d'un homme (professeur), tous impliqués en sciences de l'éducation et dans la formation des enseignants. Le Tableau 2 donne à voir des détails relatifs à ces participants. En plus des participants, un chercheur référent⁴ était présent dans chacun des groupes pour jouer le rôle de médiateur.

Tableau 2
Profil des participants de l'étude

Participant	Fonction	Âge	N ^{bre} d'années d'expérience dans la fonction	Autres informations
E1	Enseignante de français au secondaire	42	20	Enseigne en S2, S3 et S4
E2	Enseignante au primaire	25	4	Enseigne en P1, P2, P4, P5 et P6
E3	Enseignante au primaire	36	18	Enseigne en P4
E4	Enseignante au primaire	24	3	Enseigne en P4
E5	Enseignante au primaire	32	11	Enseigne en P3
C1	Doctorante	47	1	A été institutrice au primaire
C2	Professeure	31	2	
C3	Professeure	43	7	
C4	Professeur	42	3	A été instituteur au primaire et formateur d'enseignants

Note. C = chercheur; E = enseignant.

Supports utilisés dans l'étude

Deux outils ont été constitués en amont de notre étude. Premièrement, 40 enregistrements vidéo ont été recueillis auprès d'élèves de 11 et 12 ans de la Fédération Wallonie-Bruxelles, qui ont dû répondre à la consigne suivante : « Explique oralement à ton camarade ce que tu sais de cet animal »⁵.

Deuxièmement, les chercheurs référents ont créé une première version de la grille (V1) sur la base des travaux de Dolz et al. (1993), de Gregg et Mather (2002), de Colognesi et al. (2020) et des *Socles de compétences* (FW-B, 2013). Dans cette première version, les capacités langagières de Dolz et al. (1993) constituent les différentes dimensions de la grille, tandis que les compétences du programme belge fondent les sous-dimensions de celles-ci. Par ailleurs, nous avons tenté de transformer les différentes notions théoriques relatives à ces compétences et capacités au travers des différents critères. Cette grille a été utilisée pour évaluer 20 des vidéos susmentionnées. Ce premier codage a permis de supprimer deux critères redondants, d'en subdiviser un et d'en reformuler certains, ce qui a donné lieu à une deuxième version de la grille (V2).

Au sein de cette grille, une échelle de mesure à quatre niveaux (allant de zéro « non, on ne retrouve pas cette information dans le discours de l'élève » à trois « oui, on retrouve cette information de manière remarquable dans le discours de l'enfant ») est ajoutée. De plus, pour certains items, un comptage du nombre de fois où l'information est présente dans le discours de l'enfant est prévu.

Récolte des données

C'est au départ de la version 2 de la grille (voir Annexe 1) que les membres du groupe ont été invités, individuellement, à coder trois vidéos différentes. Ensuite, les participants se sont regroupés par « métier » (les enseignants et les chercheurs) avec comme consignes de discuter de leur utilisation de la grille et de négocier des ajustements pour la rendre plus efficace dans des situations pouvant être vécues dans leur cadre professionnel. Ces échanges, sur lesquels se concentre cette étude, ont pris un peu plus d'une heure et demie et ont donné lieu à deux grilles retravaillées (V3) : celle des enseignants et celle des chercheurs.

Ces versions ont ensuite été mises à l'épreuve : chaque participant a codé une autre vidéo avec les deux grilles créées et a donné son avis sur celles-ci lors d'une discussion avec l'ensemble du groupe collaboratif. Par

la suite, les différents sous-groupes se sont retrouvés pour ajuster encore la grille afin d'aboutir à une version finale (V4). Tous les échanges ont été enregistrés et retranscrits. Les différentes versions des grilles ont été conservées. Les diverses étapes explicitées ci-dessus sont reprises dans la Figure 1.

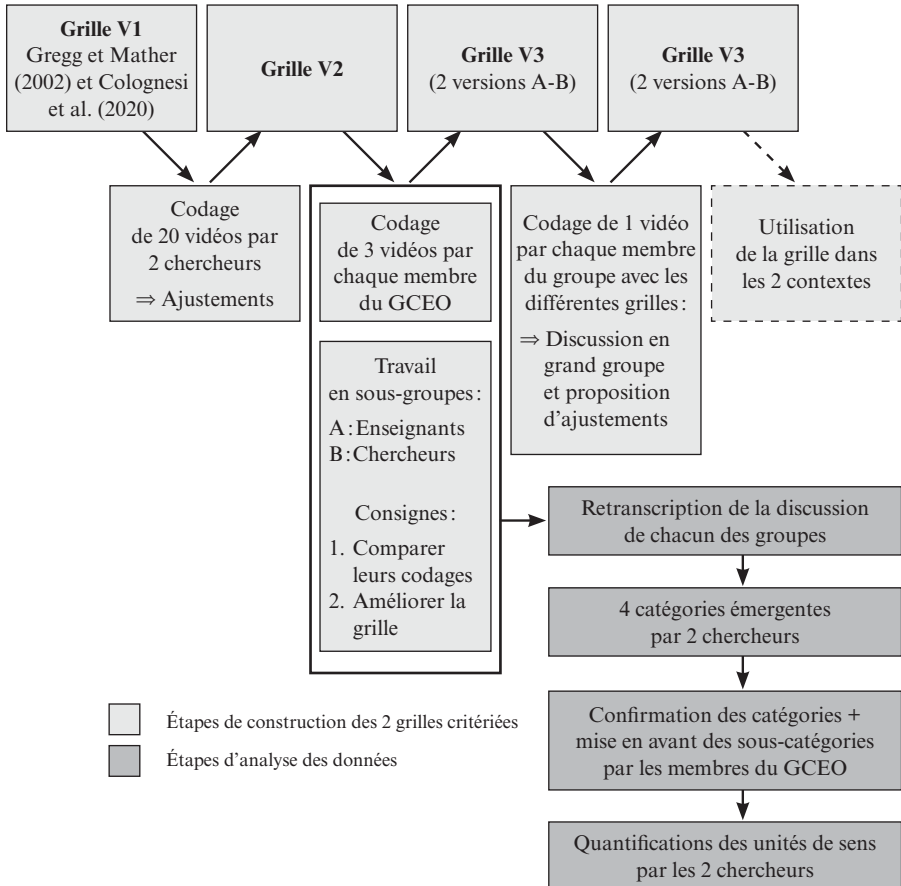


Figure 1. Synthèse du processus de coconception de la grille critériée et de l'analyse des données

Analyse des données

L'analyse des données se concentre sur les échanges réalisés en sous-groupes professionnels dans le but de construire les troisièmes versions (V3) de la grille (voir Figure 2). La retranscription de ces échanges représente un peu plus de 7000 mots pour chaque sous-groupe. Comme l'ont préconisé Bourassa et al. (2007), l'analyse des données n'a pas uniquement été menée par les chercheurs référents, mais également par l'ensemble des membres du groupe collaboratif, en suivant les prescriptions de L'Écuyer (1990) pour un travail sur un modèle ouvert.

Pour commencer, une lecture préliminaire des retranscriptions a été réalisée par les chercheurs référents, ce qui leur a permis de faire émerger quatre grandes catégories. Ensuite, ces catégories ont été présentées aux autres membres du groupe collaboratif, qui ont réalisé à leur tour une analyse de contenu sur la base des mêmes entretiens. Ainsi, les différents membres du groupe collaboratif ont pu travailler sur le même recueil de données, avec l'objectif de croiser les regards et d'ainsi augmenter la fiabilité de l'analyse.

Puis, les chercheurs référents ont pris connaissance de toutes les remarques mises en avant par les membres du groupe et ont regroupé les commentaires allant dans le même sens. Ensuite, ils se sont servis de l'ensemble de ces avis pour affiner les catégories, pour les stabiliser et pour en dégager des sous-catégories. Les membres du groupe ont également été amenés à comptabiliser les unités de sens pour chacune des sous-catégories. Par la suite, ce comptage a été retravaillé plus finement par les chercheurs référents. En tout, 483 unités de sens ont été comptabilisées. Pour synthétiser l'analyse, une schématisation d'après le modèle de Coppe et al. (2018) a été réalisée par les chercheurs référents (voir Figure 2). Elle est présentée dans la section suivante.

Résultats

La Figure 2 donne à voir la synthèse des catégories et sous-catégories résultant de l'analyse réalisée. Les grands ensembles y représentent les quatre catégories identifiées à la fois chez les enseignants et les chercheurs : (a) *Questionnements et difficultés*, (b) *Modalités à suivre pour utiliser cette grille*, (c) *Types de justifications* et (d) *Pistes d'amélioration et de solution*. Les trois premières renvoient à notre première question de recherche (les

points discutés par les acteurs), tandis que la dernière catégorie apporte des éléments de réponse à notre seconde question de recherche (les décisions et ajustements).

Dans la Figure 2, pour chacune des catégories, les rectangles correspondent aux sous-catégories établies. Plus ces rectangles sont proches du centre, plus les unités de sens classées dans cette sous-catégorie sont nombreuses. Dans chacun des rectangles, une jauge permet de comparer la représentativité de la sous-catégorie pour chaque groupe : la partie claire de la jauge correspond au groupe des enseignants, tandis que la partie foncée s'applique aux chercheurs. Les différentes catégories et sous-catégories sont présentées dans les lignes suivantes, avec des verbatim emblématiques pour les illustrer.

Ainsi, relativement à notre première question de recherche, trois grandes catégories de points de discussion ressortent des interactions : des questionnements et difficultés, des modalités envisagées pour utiliser la grille et des aspects qui permettent de justifier les choix opérés dans l'utilisation des critères.

Questionnements et difficultés

L'analyse réalisée met en évidence de manière massive (24 unités de sens pour les enseignants et 23 pour les chercheurs) la difficulté de comprendre «la signification des mots et des critères utilisés dans la grille». Les participants, qu'ils soient enseignants ou chercheurs, parlent de manière identique du manque de transparence de certains critères, de ceux qui ne sont pas assez explicites et/ou qui pourraient susciter une compréhension différente d'un utilisateur de la grille à l'autre. Cela pose la question de la convertibilité en informations de toute production orale.

Pour moi, l'item sur le nombre d'idées différentes énoncées, c'était assez flou. J'ai un peu bogué. Qu'est-ce qu'une idée? C'était assez flou. (E1)

Et il y avait l'item sur le nombre d'informations originales. Mais à partir de quand une information est-elle originale? (C3)

Ensuite, «l'utilisation de l'échelle de mesure» est également une difficulté repérée dans les discussions qui est typique de l'évaluation par compétences (Dupont et al., 2019). Les participants se questionnent sur la manière dont ils doivent interpréter les différents niveaux de l'échelle de mesure. Ils débattent aussi de l'adéquation entre le type d'échelle proposé et le critère associé. Ce type de raisonnement est plus présent chez les

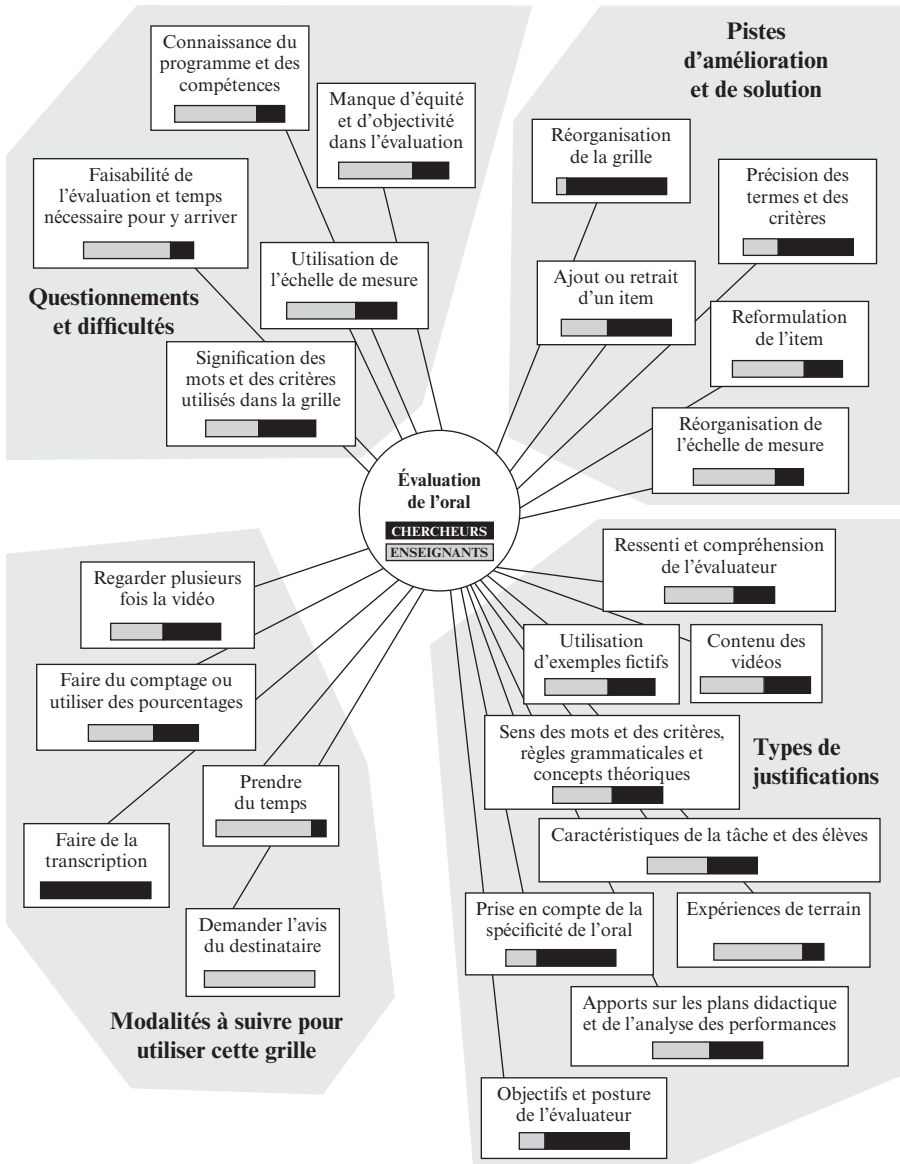


Figure 2. Schématisation des catégories et sous-catégories relatives à l'évaluation de l'oral, du point de vue des enseignants et des chercheurs

enseignants (15 unités de sens) que chez les chercheurs (6 unités de sens). Malgré cette différence, les mêmes types de réflexion ressortent dans les deux groupes.

Oui, c'est en fait des modalités 0 1 2 3. Je pense qu'il faut peut-être mettre entre «oui» et «oui de manière remarquable». Il faut peut-être ajouter [un niveau intermédiaire]. (E5)

J'ai également eu du mal à sélectionner le «1 = partiellement» et le «2 = oui». (C1)

La question de «la faisabilité de l'évaluation et du temps nécessaire pour y arriver» émerge également : les participants se demandent si la grille et les critères peuvent être utilisés facilement dans leur contexte d'utilisation. Le temps nécessaire pour utiliser la grille dans un contexte de classe est notamment au centre des discussions dans le groupe d'enseignants (11 unités de sens). Les chercheurs (3 unités de sens), quant à eux, se centrent plutôt sur les aspects annexes à la prestation de l'élève qui ont pu influencer l'évaluation réalisée.

Bien, en fait, moi, je me disais que si c'était à faire pour une classe, c'est infaisable passer 20 minutes à corriger une copie. (E3)

Si on veut bien évaluer ce critère, il faut changer la consigne donnée pour la tâche. (C4)

Les deux groupes mettent aussi en évidence que certains critères ne permettent pas une «évaluation juste et équitable de l'élève», notamment à cause de la part de subjectivité de l'évaluateur ou encore parce que les élèves qui parlent plus ont plus de risques de commettre des erreurs pour certains critères (6 unités de sens chez les enseignants et 3 chez les chercheurs). De plus, une réflexion quant aux «compétences orales» elles-mêmes a été menée par les participants des deux groupes (3 unités de sens chez les enseignants et 1 chez les chercheurs).

Parce que, dans le programme, pour les unités lexicales, il n'y a pas aussi tout ce qui est liaison et tout ça? (E2)

Je me suis dit qu'il y avait peut-être d'autres catégories en langage oral, mais je n'étais pas familière avec ça. (C2)

Modalités à suivre pour utiliser la grille

L'analyse révèle que, dans les deux groupes, les participants ont abordé ce qu'ils sont prêts à faire pour utiliser la grille le plus correctement possible ou, au contraire, ce qu'ils n'envisagent pas de faire. À une fréquence

similaire dans les deux groupes (4 unités de sens pour chacun), les participants discutent du fait de « regarder plusieurs fois la vidéo » ; ils mettent en avant la nécessité, mais également la lourdeur de cette procédure.

Non mais, je veux dire, il faut déjà regarder la vidéo plein de fois. (E1)

En me disant, si je suis chercheur, je n'ai en fait aucun problème à revoir la vidéo ou à retranscrire effectivement. On aura besoin des retranscriptions pour pouvoir compter ou mettre en fluo. (C4)

Plusieurs autres stratégies sont évoquées, dont « compter le nombre de mots » et « prendre du temps » par les deux groupes. Seuls les chercheurs abordent le fait de « retranscrire le contenu », alors que les enseignants expliquent qu'ils s'aideraient de « l'avis des autres élèves de la classe » pour compléter les items.

Par ailleurs, les enseignants ont eu tendance à réfléchir de manière plus globale en se demandant ce que leurs collègues, de manière générale, seraient prêts à faire pour utiliser la grille. Les chercheurs ont plutôt mis en évidence les stratégies qu'ils ont mises en place eux-mêmes pour réaliser la tâche. De plus, ils ont à plusieurs reprises mis en avant que, s'ils sont prêts à prendre du temps, à faire de la retranscription ou encore à regarder plusieurs fois la vidéo, c'est justement parce que leur position de chercheur et les caractéristiques qui y sont associées leur permettent/demandent de le faire.

Types de justifications

Au cours des échanges, les participants ont également justifié à plusieurs reprises la manière dont ils ont évalué les trois vidéos et précisé les raisons sous-tendant leur volonté de modifier certains critères.

Dans les deux groupes, le « ressenti et la compréhension de la situation » sont utilisés massivement. Ainsi, les participants explicitent leurs avis par ce qu'ils ont eux-mêmes vécu et pensé (32 unités de sens pour les enseignants et 25 pour les chercheurs).

Là, c'est vraiment chouette, je trouve. C'est vraiment une bonne question. (E1)

Je crois que ça ne va pas être facile. (C3)

Dans les deux groupes, les participants reprennent ce qui a été dit ou fait par les élèves pour justifier leurs propos. Ils se servent aussi du « contenu des vidéos » et des différences dans les prestations des élèves pour conforter leur raisonnement (24 unités de sens pour les enseignants et 17 pour les chercheurs).

Moi, j'avais mis 2 parce que [l'élève] le disait quand même. En commentaire, j'ai mis que ça arrivait tard, mais il en parle. J'ai mis 2 pour ça. (E3)

Je trouve qu'il y a une introduction. [L'élève] dit quand même «je parle d'un animal». C'est ça qu'il dit. (C1)

Les participants des deux groupes n'hésitent pas à inventer des «exemples fictifs» pour expliciter leur raisonnement (23 unités de sens pour les enseignants et 18 pour les chercheurs). Ils s'appuient également sur des «*éléments trouvés dans la littérature, sur des règles grammaticales ou encore sur la signification d'un mot ou d'un critère*» pour justifier leur choix (17 unités de sens pour les enseignants et 11 pour les chercheurs). Plus précisément, les enseignants ont plutôt tendance à revenir sur le sens du critère comme énoncé dans la grille, alors que les chercheurs vont plutôt se centrer sur différents concepts théoriques qui pourraient appuyer leurs propos.

Pour l'introduction, moi, je m'étais arrêté sur les mots «introduction», «introduis». Je me suis donc dit que ce n'était pas introduit. (E2)

En fait, c'est parce qu'il y a de l'intonation grammaticale et de l'intonation expressive. Il y a deux types d'intonation. (C3)

Les «caractéristiques de la tâche ainsi que des élèves» servent aussi dans les justifications. Les participants des deux groupes (13 unités de sens pour les enseignants et 10 pour les chercheurs) mettent en évidence que certains critères ne semblent pas pertinents dans le cadre de la tâche proposée, mais qu'ils pourraient l'être dans un autre contexte, d'où l'importance de les garder. Se pose ici la question de l'évaluation et de sa fonction : une seule grille peut-elle être universelle et compatible avec l'ensemble des productions orales scolaires ? Les enseignants, plus souvent que les chercheurs, ont également mis en évidence les acquis et les difficultés des élèves de fin de primaire pour justifier la pertinence de certains critères.

Avec des 6^e années, [si les élèves se trompent] 50 % [du temps], ils doivent avoir 0. Ça veut dire que ce n'est pas compréhensible. (E3)

Les phrases grammaticalement correctes et les temps corrects, je pense que c'est simplement des accords pour tout le monde. On est dans des phrases toujours à l'indicatif présent, donc je ne pense pas que ce soit un bogue. (C3)

Enfin, les enseignants (12 unités de sens), beaucoup plus que les chercheurs (1 unité de sens), expliquent leurs choix en faisant appel à leur «expérience de terrain» : ils se basent sur ce qu'ils ont déjà vécu en classe pour venir appuyer leurs propos.

En expression écrite, c'est ce qu'on fait. Moi, je compte le nombre de mots écrits, puis je fais le pourcentage. (E5)

Par exemple, dans un truc où on avait trouvé des tablettes et tout, [les élèves] misaient tellement sur leur posture, leurs gestes et tout qu'ils oubliaient de faire attention à leurs phrases, donc il faut aussi en tenir compte. (C4)

Bien que les enseignants utilisent à plusieurs reprises, tout comme dans l'extrait précédent, ce qu'ils font à l'écrit pour argumenter leur choix pour l'oral, les deux groupes vont également s'attarder sur les « spécificités de l'oral » comparativement à l'écrit pour justifier leurs choix (3 unités de sens pour les enseignants et 10 pour les chercheurs).

Moi aussi, je me suis basée sur la cohérence. Je me suis dit qu'à partir du moment que c'est fait oralement, c'est vrai que tu ne vas peut-être pas faire aussi attention sur tes mots-liens, etc. (E5)

Moi, je n'ai justement pas écrit en me disant que le risque, si j'écris, c'est de passer en évaluation de l'écrit, et pas en évaluation spontanée de l'oral. (C3)

Tous les participants utilisent également ce que la grille pourrait leur « apporter dans leur pratique » pour justifier leur raisonnement. Plus précisément, pour justifier la pertinence de certains critères de la grille, les chercheurs mettent en évidence la possibilité d'une mesure fine et approfondie des prestations des élèves grâce à ceux-ci (10 unités de sens). Les enseignants insistent plutôt sur ce que cette grille et ses critères pourraient leur apporter sur le plan didactique, notamment dans la possibilité de donner une rétroaction constructive aux élèves (9 unités de sens). Cela devient possible, à condition que ladite grille s'inscrive dans le cadre d'un dispositif d'enseignement-apprentissage et qu'elle soit suffisamment fonctionnelle pour permettre la rétroaction vers les élèves.

Et on peut discuter avec l'enfant, voilà. On peut regarder avec lui la vidéo et constater qu'il donne beaucoup d'idées, mais toutes dans la même catégorie. (E1)

Mais là, je vois assez bien le chercheur vraiment identifier les procédés utilisés par l'enfant pour élaborer ou pas les idées. (C3)

Les « objectifs poursuivis ainsi que la posture de l'évaluateur » sont aussi abordés, mais peu fréquemment (2 unités de sens chez les enseignants et 7 pour les chercheurs). Ils ont une influence certaine sur la direction prise dans la discussion des chercheurs. Ceux-ci mettent en avant qu'une de leurs fonctions en tant que chercheurs est de mesurer les performances de

l'élève, ce qui influencera leur manière d'utiliser et d'améliorer la grille. Par ailleurs, ils vont également se baser sur les projets de recherche ultérieurs pour justifier certains changements à apporter à la grille.

Oui, mais je suis sûr que si tu mets ça dans la main des profs, ils s'attendent à un *high level* et ils ne vont jamais noter 3. (E1)

Comment on veut montrer un écart qui peut-être se réalisera si on a déjà mis 3 de manière non déficitaire? Je me dis qu'ici, en tant que chercheur, si on mesure, il ne faut pas être trop «généreux». (C4)

La notion de catégorie m'intéresse plus que celle d'idée, surtout si on veut travailler après sur les connecteurs logiques. (C3)

Notre seconde question de recherche concerne les ajustements que les participants envisagent ou ont réalisés au fil de leurs échanges. Il est donc question ici des pistes d'amélioration et de solution.

Pistes d'amélioration et de solution

La plupart des unités de sens répertoriées dans cette catégorie sont des solutions proposées directement par les participants en réponse «aux questionnements et difficultés» présentés *supra*.

La proposition de «réorganiser l'échelle de mesure» est la solution la plus représentée (33 unités de sens pour les enseignants et 9 pour les chercheurs). Cette constatation est cohérente avec le fait que de nombreux questionnements ont été mis en évidence dans les deux groupes par rapport à «l'utilisation de l'échelle de mesure». D'autant plus que le nombre de modalités dans l'échelle peut influencer le positionnement.

Ce type de solution permet également de diminuer certaines difficultés relatives à la «faisabilité et à l'équité dans l'évaluation». La solution de faire varier l'étalonnage de l'échelle en fonction du critère est proposée par les deux groupes. Cependant, les enseignants décident d'inclure l'explication du code de l'échelle au sein de la grille pour avoir des repères plus concrets de ce qui est attendu pour chaque item. Dans leurs discussions, les enseignants prennent ainsi le temps, pour chaque critère, d'identifier le code qui sera utilisé. De leur côté, les chercheurs gardent un code à l'extérieur de la grille et vont passer moins de temps à changer l'échelle de mesure. Cette volonté de changer l'échelle de mesure est facilement visible dans les deux nouvelles grilles, mais est plus marquée dans la version des enseignants, pour lesquels le travail sur l'échelle de mesure est le plus grand changement apporté à la grille (voir Figure 3).

Version utilisée en début de discussion (V2)

3- Organisation générale du texte – respect de la superstructure	Spontané			
Présence de l'introduction du sujet	0	1	2	3
Les différents éléments sont introduits les uns après les autres de manière cohérente.	0	1	2	3
Présence d'une conclusion	0	1	2	3
Nombre de connecteurs utilisés adéquatement				
Nombre de connecteurs utilisés de manière non adéquate (dont tics de langage)				

Version améliorée par les enseignants (V3)

3- Organisation générale du texte – respect de la superstructure			
Présence de l'introduction du sujet	0 Pas d'introduction		2 L'élève explique de quoi il va parler.
Les différents éléments sont introduits les uns après les autres (connecteurs, mots-liens, annonce de la catégorie).	0 non		2 oui
L'élève passe d'une catégorie à l'autre sans faire d'aller-retour.	0 non		2 oui
Présence d'une formule de clôture	0 Rien	1 Un mot (ex. : voilà)	2 Une phrase
Utilisation de connecteurs	0 Pas d'utilisation	1 Utilisation de connecteurs	2 Utilisation de connecteurs variés et bien utilisés

Version améliorée par les chercheurs (V3)

3- Organisation générale du texte – respect de la superstructure	Spontané			
Présence de l'introduction du sujet	0	1		
Qualité de l'introduction	0	1	2	3
Les différents éléments sont introduits les uns après les autres de manière cohérente.	0	1	2	3
Temps d'explication de la superstructure (temps de synthèse)	0	1	2	3
Présence d'une conclusion	0		1	
Qualité de la conclusion	0	1	2	3
Nombre de connecteurs utilisés adéquatement (par catégorie)				

Figure 3. Comparaison de l'échelle de mesure dans les différentes versions de la grille

Afin de répondre aux difficultés relatives à la « signification des mots et des critères utilisés dans la grille », trois types de solutions ont été abordés par les participants. D'abord, la plus utilisée chez les enseignants (17 unités de sens contre 10 pour les chercheurs) est la « reformulation » de certains critères. Ensuite, les deux groupes vont « rajouter » certains critères afin de séparer chaque idée les unes des autres, puis d'en « supprimer » d'autres pour simplifier la grille. Cette technique semble utilisée un peu plus par les chercheurs (1 unité de sens contre 10 pour les enseignants). Enfin, « préciser » ce qu'un critère signifie, notamment au sein d'un document annexe permettant à l'évaluateur d'avoir plus d'informations sur la manière dont il doit utiliser la grille, est également plus utilisé chez les chercheurs (12 unités de sens contre 4 pour les enseignants).

La « réorganisation de la grille » est également abordée à raison d'une seule proposition d'amélioration allant dans ce sens chez les enseignants. Ceux-ci vont transférer un des items de la compétence 1 vers la compétence 2 (voir critère grisé dans la Figure 4). Quant aux chercheurs, c'est le type de solution qui a eu le plus d'impact et d'influence sur leur nouvelle version de la grille (10 unités de sens). Comme le montre la Figure 4, les chercheurs ont modifié le système de la grille, de sorte que certaines parties ne fonctionnent plus de la manière initialement prévue. Par exemple, ils vont imaginer un système en arbre qui permet au chercheur de retranscrire certains éléments du discours de l'élève afin de pouvoir analyser plus finement les productions.

Discussion et conclusion

L'évaluation spécifique de l'oral reste complexe et peu présente dans de nombreuses classes francophones (Colognesi et Deschepper, 2019; Dumais et al., 2017). Ces difficultés sont entre autres nourries par le flou conceptuel relatif à l'oral (Dumais, 2014; Dumais et Lafontaine, 2011) et par le peu d'outils existants pour guider les enseignants dans cette démarche (Alrabadi, 2011; Lafontaine et Messier, 2009). Ce manque d'outils fidèles et valides peut également compliquer la mise en place d'études permettant d'en apprendre plus sur le développement et sur le niveau actuel des élèves dans les différentes composantes de l'oral (Nonnon, 2016). Devant ces difficultés éprouvées tant du côté des enseignants que des chercheurs, nous avons étudié quels seraient leurs points de discussion s'ils étaient amenés à évaluer des prestations orales d'élèves du primaire à l'aide d'une grille

Version utilisée en début de discussion (V2)

B. Élaborer des contenus/des significations				
L'ensemble du discours reste centré sur le sujet ou le nourrit.	0	1	2	3
Les idées sont élaborées (accompagnées d'exemple, etc.).	0	1	2	3
Nombre d'idées différentes énoncées				
Nombre de catégories abordées				
Nombre d'informations erronées énoncées				
Nombre d'informations originales (utilisation de métaphore, d'image, de contenu spécifique, de comparaison, etc.)				

Version améliorée par les enseignants (V3)

B. Élaborer des contenus/des significations			
L'explication permet la compréhension de l'objet par l'orateur.	0 non		2 oui
L'ensemble du discours reste centré sur le sujet.	0 non		2 oui
Les idées sont développées (présence d'adjectifs, d'idées plus précises, de métaphores, d'illustrations et/ou de comparaisons).	0 Aucune des idées n'est développée	1 Moins de la moitié des idées sont développées	0 Plus de la moitié des idées sont développées
Nombre d'informations différentes			
Nombre de catégories abordées			
Nombre d'informations erronées			

Version améliorée par les chercheurs (V3)

B – Élaborer des contenus/des significations – Arbre								Total
Catégories	1.	2.	3.	4.	5.	6.	7.	
Idées globales								
Idées originales								
Idées inutiles								

Ligne du temps								
0								

Figure 4. Comparaison de l'organisation de la grille dans ses différentes versions

critériée, et quelles seraient les décisions prises pour élaborer une grille au plus proche de leur terrain professionnel. Nous avons également mis en perspective ces aspects en les comparant d'un groupe à l'autre. L'objectif à long terme de cette démarche est de créer des outils d'évaluation adaptés au monde de l'enseignement et de la recherche pour mesurer au plus près les compétences orales des élèves.

Ressemblances entre les deux groupes

De nombreuses ressemblances ont été mises en évidence entre le groupe des enseignants et des chercheurs. Effectivement, 22 sous-catégories sur les 24 répertoriées sont conjointes aux deux groupes, ce qui amène à penser que les aspects discutés sont relativement similaires. Par exemple, et de manière forte, les deux groupes mettent en évidence des difficultés à «comprendre la signification des mots utilisés dans la grille» et à se mettre d'accord sur le sens des critères. Les enseignants n'étant pas toujours informés des différents référents théoriques liés à l'oral, cette difficulté ne semble pas étonnante pour ce groupe. À l'inverse, il est plus surprenant que des chercheurs travaillant dans le domaine de l'oral éprouvent également ces difficultés puisqu'ils sont, nous pouvons le penser, plus au fait des modèles didactiques utilisés pour construire la grille. Cependant, ces modèles ont été transformés en critères et c'est dans ce processus de reformulation que les chercheurs impliqués ne sont pas toujours d'accord. Ainsi, les résultats tendent à mettre en évidence que le manque de clarté dans la formulation des critères, qui est un des obstacles principaux à l'utilisation des grilles critériées (Baume et al., 2004), est valable non seulement pour les enseignants, mais aussi pour les chercheurs. Par ailleurs, ce qui semble marquer l'intérêt, lors de la conception d'une grille critériée, c'est de faire tester l'outil auprès d'un échantillon de futurs utilisateurs pour s'assurer que les critères sont compréhensibles et utilisables.

Le fait que, dans les deux groupes, les participants semblent suivre une régulation épistémique où chacun tente d'intégrer le point de vue de l'autre dans la conception de la grille (Darnon et al., 2006) est également une ressemblance que notre étude met en évidence.

Différences entre les deux groupes

Plusieurs différences ont aussi été identifiées entre les deux groupes. En effet, à la suite de Morrissette et al. (2012), les enseignants de notre étude ont fait régulièrement référence, dans leur discussion, au contexte

de classe : ils abordent, plus souvent que nos chercheurs, les difficultés concrètes liées à l'utilisation de la grille. Également, plus que les chercheurs, ils prennent en considération leurs collègues et ce qu'ils seraient prêts à faire ou non pour évaluer l'oral ainsi que les normes de référence des enseignants, normes qui, comme le mettent en évidence Dupont et Grandaty (2018), sont influencées par les situations langagières et par les références de l'écrit.

Par ailleurs, pour justifier leur point de vue, les cinq enseignants, de manière générale, se basent sur des éléments plus concrets, par exemple les expériences qu'ils ont eux-mêmes vécues dans leur classe.

Les enseignants ont également directement mis en évidence les apports de la grille en matière de création de séquences didactiques et de rétroaction qui pourraient être données aux élèves (Balan et Jönsson, 2018 ; Berdal-Masuy et Renard, 2015 ; Berthiaume et al., 2011).

Les chercheurs de notre étude se sont plutôt concentrés sur la nécessité d'obtenir une mesure de l'oral précise et de qualité à l'aide de la grille. Ils ont également mis en évidence l'importance de pouvoir, grâce à la grille, observer finement l'évolution des élèves lors de différents temps de mesure au sein d'une intervention dans les classes. Dans leurs justifications, ils ont insisté plusieurs fois sur le fait qu'en tant que chercheurs, ils suivent une posture d'évaluateur, et non de formateur, ce qui implique une certaine rigueur dans l'utilisation de la grille.

Les chercheurs justifient également quelques prises de position par la nécessité pour eux d'obtenir des données suffisantes pour réaliser des projets de recherche, par exemple celui d'en apprendre plus sur l'utilisation des connecteurs par des élèves de fin de primaire. Globalement, ils se sont moins concentrés sur le ressenti des élèves et les apports directs de la grille sur ceux-ci. Contrairement aux enseignants, la question du temps nécessaire pour réaliser le codage ne semble pas un obstacle.

Au final, le cadre de référence de chacun, créé à partir des expériences et des contraintes du métier (Coen et al., 2008 ; Coen et Pellegrini, 2011), vient influencer l'orientation de la discussion, mais aussi les ajustements qui sont apportés dans la grille pour la perfectionner en conséquence.

Le groupe d'enseignants a gardé le cap d'une utilisation peu chronophage. Il a pris le temps de discuter de l'échelle, de la simplifier, de l'adapter à chaque critère et de spécifier la signification de chaque échelon. Ces procédures permettent, selon Dunbar et al. (2006), un codage plus similaire

entre les différents utilisateurs de la grille. Le groupe des enseignants a également reformulé de nombreux items afin qu'ils soient plus explicites et compréhensibles pour n'importe quel enseignant, que celui-ci soit expert ou non, palliant ainsi un des problèmes liés aux grilles critériées (Balan et Jönsson, 2018 ; Baume et al., 2004).

Le groupe de chercheurs, quant à lui, s'est servi des cadres théoriques issus des sciences de l'éducation et de la didactique du français pour affiner la grille et la rendre plus précise (Morrissette et al., 2012), notamment dans l'idée que l'oral ne peut être considéré comme l'écrit, par exemple en matière de règles grammaticales et syntaxiques (Alrabadi, 2011 ; Nonnon 2016). Il s'assure ainsi que les critères correspondent à ce qu'il est censé mesurer. En plus, ce groupe ajoute un document annexe à la grille pour guider les évaluateurs dans l'utilisation de la grille, dans le but de tenter d'augmenter la fidélité interjuges de cet outil.

Nous pouvons constater que le travail en groupe de pairs a permis l'amélioration d'une grille critériée qui, grâce à de nombreuses discussions, intègre le cadre de référence de chacun des membres. De plus, les pistes d'amélioration proposées par les deux groupes sont distinctes, ce qui permet la création de deux grilles différentes. Ainsi, interroger l'expertise professionnelle de ces deux corps de métiers semble une piste intéressante dans la conception d'outils d'évaluation plus en adéquation avec les différentes réalités professionnelles au sein desquelles ils seraient utilisés.

Limites

Cependant, certaines limites sont à pointer. Premièrement, un seul groupe de chercheurs et un seul groupe d'enseignants ont été sollicités pour réaliser cette étude, et les participants sont tous des membres d'un même groupe collaboratif. Ainsi, nous pouvons penser que les points de discussion et les ajustements pointés dans les groupes auraient pu varier avec d'autres participants. Pour pallier cette limite, cette grille sera testée, dans une étape ultérieure, auprès d'enseignants et de chercheurs extérieurs au projet afin de confirmer les changements réalisés et d'ajouter des ajustements éventuels.

Deuxièmement, pour cette étude, les participants utilisaient la grille pour évaluer une prestation orale qui n'est pas issue d'une séquence d'apprentissage. En effet, il a été demandé aux élèves de prendre la parole sans qu'il y ait eu d'enseignement préalable. Cette démarche, réalisée dans un contexte de recherche, visait à en apprendre plus sur les prérequis des

élèves en matière de compétences orales. Partant, elle permettra d'effectuer un état des lieux sur les besoins des élèves relativement à l'explication orale et de développer des séquences d'enseignement-apprentissage en fonction. Dans ce cadre, la grille pourrait être mobilisée dans ce contexte authentique, de manière à devenir un outil d'évaluation formative à destination des apprenants. Plus encore, dans une perspective de « réoralisations » (Colognesi et al., 2017), elle apportera des informations aux élèves et à l'enseignant pour pouvoir améliorer leur communication.

Troisièmement, cette étude s'est centrée uniquement sur les enseignants et les chercheurs. Or, dans un contexte de classe, les élèves sont également des acteurs essentiels de ce processus d'évaluation. D'ailleurs, certains auteurs, comme Campanale (2015), mettent en évidence qu'il est pertinent de les inclure dans le processus de conception des outils d'évaluation. C'est ce à quoi nous nous attelons actuellement. Ainsi, transposer la démarche utilisée dans cette étude avec un groupe d'élèves permettrait de mettre en évidence d'autres points de discussion et d'autres types de solutions spécifiques, de sorte que l'outil soit adapté aux besoins et attentes de ce public.

Enfin, un temps conséquent est nécessaire pour remplir ces nouvelles grilles critériées, ce qui n'est pas toujours compatible avec un contexte de classe. Par ailleurs, arriver à mesurer toutes les composantes de la compétence à communiquer oralement uniquement par le biais de critères reste un défi majeur (Bouwer et al., s. d. ; Derycke, 2000). Dans le même ordre d'idées, comme les participants des deux groupes le mettent en évidence, il n'est pas toujours aisé de transformer des éléments d'une production orale en une information mesurable dans une grille. Au vu de ces difficultés intrinsèques aux grilles critériées, travailler avec des outils d'évaluation autres que les grilles critériées dans le but de les compléter semble une piste d'exploration pertinente pour s'adapter le plus possible aux enseignements de l'oral réalisés en classe (Dupont, 2020).

Réception : 13 juillet 2020

Version finale : 3 mai 2021

Acceptation : 4 mai 2021

NOTES

1. Le *savoir parler* et le *savoir écouter* sont des termes repris dans les *Socles de compétences* (Fédération Wallonie-Bruxelles, 2013) belges et font référence à la face expressive et réceptive de l'oral.
2. Le terme «oral» comprend aussi bien la prise de parole que la compréhension. Dans notre étude, seule la prise de parole a été mobilisée, ce qui correspond à la «compétence à communiquer oralement», d'après Dumais et Soucy (2020). En référence à ces auteurs, c'est cette terminologie que nous utiliserons dans la suite de ce texte.
3. En Belgique, l'enseignement obligatoire est divisé en deux niveaux successifs : (a) l'enseignement primaire, allant de la 1^{re} année à la 6^e année, et accueillant des enfants de 6 à 12 ans et (b) l'enseignement secondaire, allant de la 1^{re} secondaire jusqu'à la 6^e ou 7^e année, en fonction des options choisies. Les enseignants du primaire peuvent soit être titulaire de leur classe et ne travailler que dans celle-ci, soit être enseignant volant et enseigner dans diverses classes de différentes années.
4. Les chercheurs qui ont créé et organisé l'étude ne sont pas les mêmes que les participants membres du sous-groupe chercheur. Les chercheurs organisant l'étude seront nommés «chercheurs référents» dans la suite de cet article afin de les différencier des chercheurs ayant un rôle de participant.
5. Les vidéos utilisées comme support dans notre étude ont été récoltées dans l'optique de mesurer les prérequis des élèves dans les différentes composantes de l'oral sans enseignement préalable. Ainsi, aucune séquence d'apprentissage ne précède l'enregistrement de ces vidéos.

RÉFÉRENCES

- Allen, N. (2017). Développer la littératie volet oral pour favoriser la transition interordre entre le primaire et le secondaire : quelques considérations théoriques. *Canadian Journal of Education*, 40(4), 638-664. <https://journals.sfu.ca/cje/index.php/cje-rce/article/view/3115>
- Alrabadi, E. (2011). ¿Qué método se debe adoptar para la enseñanza/aprendizaje de la comunicación oral? [Quelle méthodologie faut-il adopter pour l'enseignement-apprentissage de l'oral?] *Didáctica. Lengua y Literatura*, 23, 15-34. https://doi.org/10.5209/rev_DIDA.2011.v23.36308
- Balan, A. et Jönsson, A. (2018). Increased explicitness of assessment criteria : Effects on student motivation and performance. *Frontiers in Education*, 81(3). <https://doi.org/10.3389/educ.2018.00081>
- Baume, D., Yorke, M. et Coffey, M. (2004). What is happening when we assess, and how can we use our understanding of this to improve assessment? *Assessment & Evaluation in Higher Education*, 29, 451-477. <https://doi.org/10.1080/02602930310001689037>

- Bednarz, N., Desgagné, S. et Savoie-Zajc, L. (2012). La mise au jour d'un contrat réflexif comme régulateur de démarches de recherche participative : le cas d'une recherche-action et d'une recherche collaborative. *Recherches en éducation*, 14, 128-151. <http://dx.doi.org/10.4000/rec.5894>
- Berdal-Masuy, F. et Renard, C. (2015). Comment évaluer l'impact des pratiques théâtrales sur les progrès en langue cible? Vers un nouveau dispositif d'évaluation de l'oral en FLE. *Lidil*, 52, 153-174. <https://doi.org/10.4000/lidil.3872>
- Berger, J.-L. et Büchel, F. P. (2013). *L'autorégulation de l'apprentissage : perspectives théoriques et applications*. Éditions Ovidia.
- Berthiaume, D., David, J. et David, T. (2011). Réduire la subjectivité lors de l'évaluation des apprentissages à l'aide d'une grille critériée : repères théoriques et applications à un enseignement interdisciplinaire. *Revue internationale de pédagogie de l'enseignement supérieur*, 27(2). <http://ripes.revues.org/524>
- Biggs, J. B. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347-364. <https://doi.org/10.1007/BF00138871>
- Bourassa, M., Bélair, L. et Chevalier, J. (2007). Liminaire : les outils de la recherche participative. *Éducation et francophonie*, 35(2), 1-11. https://revue.acef.ca/pdf/XXXV_2_000.pdf
- Bouwer, R., Koster, M. et Van den Bergh, H. (s. d.). *Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner*. Document soumis pour publication en 2016.
- Buchs, C., Darnon, C., Quiamzade, A., Mugny, G. et Butera, F. (2008). Régulation des conflits sociocognitifs et apprentissage. *Revue française de pédagogie*, 163, 105-125. <https://doi.org/10.4000/rfp.1013>
- Campanale, F. (2005, 24-26 octobre). *La grille de critères pour autoévaluer : norme à respecter ou outil évolutif pour progresser?* Communication orale présentée au 18^e Colloque international de l'ADMEE-Europe, Reims, France. <https://hal.archives-ouvertes.fr/hal-01175415/file/FrancoiseCampanalecomadmee05grilles%20de%20criteres-bis.pdf>
- Chabanne, J. et Bucheton, D. (2002). *Parler et écrire pour penser, apprendre et se construire : l'écrit et l'oral réflexifs*. PUF.
- Coen, P.-F., Galland, A., Monnard, I., Perrin, N. et Rouiller, Y. (2008). Pertinence et validité des tâches complexes dans l'évaluation des futurs enseignants. Dans L. Paquay, G. Baillat, J. De Ketele et C. Thélot (dir.), *Évaluer pour former : outils, dispositifs, et acteurs* (p. 189-204). De Boeck.
- Coen, P.-F. et Pellegrini, S. (2011, 12-14 janvier). *Quand plusieurs correcteurs évaluent les mêmes tâches complexes : regards sur la concordance inter-juges et point de vue des évaluateurs*. Communication orale présentée au 23^e Colloque de l'ADMEE-Europe, Paris, France. http://www.sr-hepfr.ch/taco/taco/contributions_files/Colloque%20ADMEE.pdf
- Colognesi, S. et Deschepper, C. (2019). Les pratiques déclarées de l'enseignement et de l'évaluation de l'oral au primaire : qu'en est-il en Belgique francophone? *Language & Literacy*, 21(1), 1-18. <https://doi.org/10.20360/langandlit29365>
- Colognesi, S. et Dolz-Mestre, J. (2017). Faire construire des scénarios pour développer les capacités orales des élèves du primaire. Dans J.-F. de Pietro, Carole Fisher et R. Gagnon (dir.), *L'oral aujourd'hui : perspectives didactiques* (p. 177-196). Presses universitaires de Namur. <https://archive-ouverte.unige.ch/unige:97894>

- Colognesi, S. et Hanin, V. (2020). Quelles pratiques efficaces pour enseigner l'oral? Expérimentations dans 8 classes du primaire et suivi de 16 futurs enseignants. *Recherches*, 73(1). https://www.researchgate.net/publication/337917836_Quelles_pratiques_efficaces_pour_enseigner_l'oral_Experimentations_dans_8_classes_du_primaire_et_suivi_de_16_futurs_enseignants
- Colognesi, S., Lyon López, N. et Deschepper, C. (2017). Développer l'enseignement de l'oral au début du primaire: l'impact de deux activités spécifiques en parler-écouter. *Language & Literacy*, 19(4), 55-75. https://www.researchgate.net/publication/320050610_Le_developpement_de_l'oral_a_l'ecole_fondamentale_l'impact_de_deux_activites_specifiques_en_parler-ecouter
- Colognesi, S., Vassart, C., Blondeau, B. et Coertjens, L. (2020). Formative peer assessment to enhance primary school pupils' oral skills: Comparison of written feedback without discussion or oral feedback during a discussion. *Studies in Educational Evaluation*, 67, 100917. <https://doi.org/10.1016/j.stueduc.2020.100917>
- Coppe, T., März, V., Decuyper, M., Springuel, F. et Colognesi, S. (2018). Ouvrir la boîte noire du travail de préparation de l'enseignant : essai de modélisation et d'illustration autour du choix et de l'évolution d'un document support de cours. *Revue française de pédagogie*, 3(204), 17-31. <https://doi.org/10.4000/rfp.8358>
- Darnon, C., Muller, D., Schragger, S. M., Pannuzzo, N. et Butera, F. (2006). Mastery and performance goals predict epistemic and relational conflict regulation. *Journal of Educational Psychology*, 98, 766-776. <https://doi.org/10.1037/0022-0663.98.4.766>
- Derycke, M. (2000). La grille critériée et le portfolio à l'épreuve du suivi pédagogique. *Revue française de pédagogie*, 132, 23-32. <https://doi.org/10.3406/rfp.2000.1030>
- Desgagné, S. et Larouche, H. (2010). Quand la collaboration de recherche sert la légitimation d'un savoir d'expérience. *Recherches en éducation, hors série 1*, 7-18. https://crises.ulaval.ca/sites/default/files/fulltext/quand_la_collaboration_de_recherche_sert_la_legitimation_dun_savoir_dexperience_.pdf
- Dolz, J. et Gagnon, R. (2008). Le genre du texte, un outil didactique pour développer le langage oral et écrit. *Pratiques*, 137/138, 179-198. <https://doi.org/10.4000/pratiques.1159>
- Dolz, J., Pasquier, A. et Bronckart, J.-P. (1993). L'acquisition des discours : émergence d'une compétence ou apprentissage de capacités langagières diverses? *Études de linguistique appliquée*, 92, 23-37.
- Dolz, J. et Schneuwly, B. (1998). *Pour un enseignement de l'oral: initiation aux genres formels à l'école*. ESF Éditeur.
- Dumais, C. (2014). *Taxonomie du développement de la langue orale et typologie: fondements pour l'élaboration d'une progression des objets d'enseignement/l'apprentissage de l'oral en classe de français langue première qui s'appuie sur le développement intégral des élèves de 6 à 17 ans* [Thèse de doctorat non publiée]. Université du Québec en Outaouais.
- Dumais, C. (2016). Proposition d'une typologie des objets d'enseignement/l'apprentissage de l'oral. *Les dossiers des sciences de l'éducation*, 36, 37-56. <https://doi.org/10.4000/dse.1347>
- Dumais, C. et Lafontaine, L. (2011). L'oral à l'école québécoise : portrait des recherches actuelles. *Revue suisse des sciences de l'éducation*, 33(2), 285-302. <https://sjer.ch/article/view/4861/7136>

- Dumais, C., Lafontaine, L. et Pharand, J. (2015). Enseigner et évaluer l'oral en milieu défavorisé au Québec : premiers résultats d'une recherche-action-formation au 3^e cycle du primaire. *Language & Literacy*, 17(3), 5-27. <https://doi.org/10.20360/G2M596>
- Dumais, C., Lafontaine, L. et Pharand, J. (2017). «J'enseigne et j'évalue l'oral» : pratiques effectives au 3^e cycle du primaire. Dans J.-F. de Pietro, C. Fisher et R. Gagnon (dir.), *L'oral aujourd'hui : perspectives didactiques* (p. 151-174). Presses universitaires de Namur.
- Dumais, C. et Soucy, E. (2020). Des documents de référence pour l'enseignement et l'évaluation d'objets de l'oral au primaire québécois. *Revue hybride de l'éducation*, 4(4), 24-60. <https://doi.org/10.1522/rhe.v4i4.1068>
- Dunbar, N. E., Brooks, C. F. et Kubicka-Miller, T. (2006). Oral communication skills in higher education : Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, 31(2), 115-128. <https://doi.org/10.1007/s10755-006-9012-x>
- Dupont, P. (2019). Vitaliser l'enseignement de l'oral : la séquence didactique du genre scolaire disciplinaire. Dans K. Sénéchal, C. Dumais et R. Bergeron (dir.), *Didactique de l'oral : de la recherche à la classe, de la classe à la recherche* (p. 43-68). Peisaj.
- Dupont, P. (2020, 22-24 janvier). Un dispositif innovant pour enseigner l'oral en intégrant sa dimension évaluative, le projet SEMO. Dans S. Colognesi, P. Dupont et J. Dolz (dir.), *Quelles avancées sur l'évaluation de l'oral, pour la recherche et pour la pratique?* [communication orale]. 32^e Colloque de l'ADMEE-Europe, Casablanca, Maroc.
- Dupont, P., Buznic-Bourgeacq, P. et Carnus, M.-F. (2019). *Compétence(s) et savoir(s) pour enseigner et pour apprendre : controverses, compromis ou compromissions?* L'Harmattan.
- Dupont, P. et Grandaty, M. (2016). De la dichotomie orale enseigné-oral pour apprendre à la dialectique oral travaillé-oral enseigné. *Repères*, 54, 7-16. <https://doi.org/10.4000/repere.1071>
- Dupont, P. et Grandaty, M. (2018). L'oral, un serpent de mer? Dans S. Aeby Daghé et M.C. Guernier (dir.), *Contextes institutionnels, réformes et recherche en didactique du français* (p. 97-131). Presses universitaires de Namur.
- Fédération Wallonie-Bruxelles (FW-B). (2013). *Socles de compétences*. Bruxelles : Administration générale de l'enseignement et de la recherche scientifique, Service général du pilotage du système éducatif. <http://enseignement.be/index.php?page=24737>
- Gagnon, R., de Pietro, J.-F. et Fisher, C. (2017). Introduction. Dans J.-F. de Pietro, C. Fisher et R. Gagnon (dir.), *L'oral aujourd'hui : perspectives didactiques* (p. 11-40). Presses universitaires de Namur. <https://books.openedition.org/pun/6537>
- Gan, Z., Davidson, C. et Hamp-Lyon, L. (2009). Topic negotiation in peer group oral assessment situations : A conversation analytic approach. *Applied Linguistics*, 30(3), 315-334. <https://doi.org/10.1093/applin/amn035>
- Garcia-Debanç, C. (1999). Évaluer l'oral. *Pratiques*, 103-104, 193-212. <https://doi.org/10.3406/prati.1999.1867>
- Garcia-Debanç, C. et Delcambre, I. (2001). Enseigner l'oral? *Repères : recherches en didactique du français langue maternelle*, 24-25, 3-21. <https://doi.org/10.3406/repere.2001.2367>

- Gregg, N. et Mather, N. (2002). School is fun at recess : Informal analyses of written language for students with learning disabilities. *Journal of Learning Disabilities*, 35(1), 7-22. <https://doi.org/10.1177/002221940203500102>
- Jorro, A. (2009). L'évaluation comme savoir professionnel. Dans L. Mottier Lopez et M. Crahay (dir.), *Évaluation en tension : entre la régulation des apprentissages et le pilotage des systèmes* (p. 219-231). De Boeck.
- Lafontaine, L. (2016). Pratiques d'enseignement de l'oral au primaire au Québec : des avancées dans le domaine de la littératie. *Le français aujourd'hui*, 195(4), 37-46. <https://doi.org/10.3917/lfa.195.0037>
- Lafontaine, L. et Messier, G. (2009). Représentations de l'enseignement et de l'évaluation de l'oral chez des enseignants et des élèves du secondaire en français langue d'enseignement. *Revue du Nouvel-Ontario*, 34, 119-144. <https://doi.org/10.7202/038722ar>
- Lafontaine, L. et Préfontaine, C. (2007). Modèle didactique descriptif de la production orale en classe de français langue première au secondaire. *Revue des sciences de l'éducation*, 33(1), 47-66. <https://doi.org/10.7202/016188ar>
- L'Écuyer, R. (1990). *Méthodologie de l'analyse développementale de contenu : méthode GPS et concept de soi*. PUQ.
- Maurer, B. (2001). *Une didactique de l'oral : du primaire au lycée*. Éditions Bertrand-Lacoste.
- Morrisette, J., Mottier Lopez, L. et Tessaro, W. (2012). La production de savoirs négociés dans deux recherches collaboratives sur l'évaluation formative. Dans L. Mottier Lopez et G. Figari (dir.), *Modélisations de l'évaluation en éducation* (p. 27-43). De Boeck.
- Mottier Lopez, L. (2015a). Au cœur du développement professionnel des enseignants, la conscientisation critique : exemple d'une recherche collaborative sur l'évaluation formative à l'école primaire genevoise. *Carrefours de l'éducation*, 39(1), 119-135. <https://doi.org/10.3917/cdle.039.0119>
- Mottier Lopez, L. (2015b). Évaluation-régulation interactive : étude des structures de participation guidée entre enseignant et élèves dans le problème mathématique « Enclos de la chèvre ». *Mesure et évaluation en éducation*, 38(1), 89-120. <https://doi.org/10.7202/1036552ar>
- Mottier Lopez, L., Tessaro, W., Dechamboux, L. et Morales Villabona, F. (2012). La modération sociale : un dispositif soutenant l'émergence de savoirs négociés sur l'évaluation certificative des apprentissages des élèves. *Questions vives*, 6(18), p. 159-175. <http://questionsvives.revues.org/1235>
- Nonnon, É. (2016). 40 ans de discours sur l'enseignement de l'oral : la didactique face à ses questions. *Pratiques*, 169-170. <https://doi.org/10.4000/pratiques.3115>
- Perrenoud, P. (2000, 7-9 juin). *L'approche par compétences : une réponse à l'échec scolaire ?* [communication orale]. 20^e Colloque de l'Association québécoise de pédagogie collégiale, Montréal, Canada. https://cdc.qc.ca/actes_aqpc/2000/7E03-Perrenoud.pdf
- Pitts, J., Coles, C., Thomas, P. et Smith, F. (2002) Enhancing reliability in portfolio assessment : Discussion between assessors. *Medical Teacher*, 24, 197-201. <http://dx.doi.org/10.1080/01421590220125321>
- Plane, S. (2015, 31 août). Pourquoi l'oral doit-il être enseigné? *Cahiers pédagogiques*. <https://www.cahiers-pedagogiques.com/pourquoi-l-oral-doit-il-etre-enseigne>

- Pomplun, M., Capps, L. et Sundbye, N. (1998). Criteria teachers use to score performance items. *Educational Assessment*, 5(2), 95-110. https://doi.org/10.1207/s15326977ea0502_2
- Schneuwly, B. (1997). Vers une didactique du français oral? *Enjeux*, 39/40, 3-11. <https://archive-ouverte.unige.ch/unige:34114>
- Sweller, J., Ayres, P. et Kalyuga, S. (2011). *Cognitive load*. Springer.
- Tricot, A., Plégat-Soutjis, F., Camps, J.-F., Amiel, A., Lutz, G. et Morcillo, A. (2003). Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH. *Environnements informatiques pour l'apprentissage humain*, 391-402. <https://edutice.archives-ouvertes.fr/edutice-00000154>
- Van Nieuwenhoven, C. et Colognesi, S. (2015). Une recherche collaborative sur l'accompagnement des futurs instituteurs : un levier de développement professionnel pour les maîtres de stage. *e-JIREF*, 1(2), 103-121. https://dial.uclouvain.be/pr/boreal/object/boreal:165239/datastream/PDF_01/view
- Vinatier I. (2010). L'entretien de co-explicitation entre chercheur et enseignants : une voie d'émergence et d'expression du «sujet capable». *Recherches en éducation, hors série 1*, 111-129.
- Vygotski, L. S. (1934/2013). *Pensée et langage*. La Dispute.

Annexe 1 : Évaluation des compétences orales : version 2 d'une grille en conception

Temps : _____ secondes

Code	3	Oui de manière remarquable
	2	Oui
	1	Partiellement
	0	Non

A. Capacité d'action

1. Tenir compte des paramètres de la situation de communication					Commentaire(s)
L'explication est adaptée au destinataire (âge).	0	1	2	3	
L'explication permet la compréhension de l'objet par l'interlocuteur.	0	1	2	3	

B. Capacité discursive

2. Élaborer des contenus/des significations					Commentaire(s)
L'ensemble du discours reste centré sur le sujet ou le nourrit.	0	1	2	3	
Les idées sont élaborées (accompagnées d'exemple, etc.).	0	1	2	3	
Nombre d'idées différentes énoncées					
Nombre de catégories abordées					
Nombre d'informations erronées énoncées					
Nombre d'informations originales (utilisation de métaphores, d'images, de contenu spécifique, de comparaisons, etc.)					
3. Organisation générale du texte – respect de la superstructure					Commentaire(s)
Présence de l'introduction du sujet	0	1	2	3	
Les différents éléments sont introduits les uns après les autres de manière cohérente.	0	1	2	3	
Présence d'une conclusion	0	1	2	3	
Nombre de connecteurs utilisés adéquatement					
Nombre de connecteurs utilisés de manière non adéquate (dont tics de langage)					

C. Capacité linguistico-discursive

4a. Unités lexicales					Commentaire(s)
Nombre de mots spécifiquement en lien avec les caractéristiques de l'objet					
4b. Unités grammaticales					Commentaire(s)
Les phrases sont grammaticalement correctes (à l'oral).	0	1	2	3	
Les temps utilisés sont corrects.	0	1	2	3	
Le discours est dépourvu de faux départs.	0	1	2	3	

D. Dimension paraverbale et non verbale

5a. Oralement					Commentaire(s)			
Volume	Inaudible			Audible				
	0	1	2	3				
Débit de parole	Trop ou trop peu de pauses, discours trop rapide ou trop lent			Débit adapté, pauses appropriées				
	0	1	2	3				
Articulation	Les sons sont indistincts.			Les sons sont très distincts.				
	0	1	2	3				
Intonation et spontanéité des phrases	Monotonie du discours			Phrases expressives et diverses				
	0	1	2	3				
5b. Corporellement				Spontané	Commentaire(s)			
Regard vers l'interlocuteur en parlant				0	1	2	3	
Présence de gestes				0	1	2	3	
Absence de gestes parasites (se tortiller, etc.)				0	1	2	3	

E. Parasites à la communication

	Spontané	Commentaire(s)
Nombre d'interventions du chercheur		
Nombre d'interventions de l'interlocuteur		