



HAL
open science

Approches et validité psycho-didactiques des évaluations

Marc Vantourout, Rémi Goasdoué

► **To cite this version:**

Marc Vantourout, Rémi Goasdoué. Approches et validité psycho-didactiques des évaluations. *Éducation & formations*, 2014, e-302. halshs-01239551

HAL Id: halshs-01239551

<https://shs.hal.science/halshs-01239551>

Submitted on 7 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Approches et validité psycho-didactiques des évaluations

Marc Vantourout, Rémi Goasdoué

Laboratoire EDA
Université Paris-Descartes, Sorbonne Paris-Cité
45 rue des Saints-Pères
75006 PARIS
marc.vantourout@parisdescartes.fr
remi.goasdoue@parisdescartes.fr

RÉSUMÉ. Les débats sur la validité des évaluations scolaires, imprégnés de critères issus de la mesure, sont confinés à la validité des scores ou des notes. Ce texte, consacré à ce que nous appelons « approches et validité psycho-didactiques des évaluations » (APDE et VPD), entend renouveler ces débats. La thèse défendue est la suivante : les didactiques disciplinaires et la psychologie cognitive disposent d'éléments qui constituent des ressources – concepts, notions, résultats, méthodes, etc. – pouvant contribuer au développement d'évaluations qui se caractérisent par la qualité du diagnostic qu'elles permettent d'obtenir. Deux axes, tous deux indispensables et non substituables, permettent de définir VPD : un premier centré sur les contenus impliqués, un deuxième qui renvoie à ce que les psychologues et ergonomes appellent l'« analyse de la tâche ». Cette nouvelle forme de validité se démarque de celles qui existent car elle repose sur une analyse qualitative du processus de réponse de l'évalué et prend en compte les spécificités des apprentissages en jeu. La portée des APDE et la VPD ne se limite ni à une forme d'évaluation, ni à une discipline. L'entrée par l'analyse des épreuves et de l'activité de réponse permet au contraire de dépasser les clivages usuels entre types d'épreuves ou entre finalités de l'évaluation que l'on trouve dans les travaux consacrés à l'évaluation.

MOTS-CLÉS : évaluations – validité – diagnostic – didactiques – psychologie cognitive

1. Mesure et évaluations scolaires

Le constat des divergences entre évaluateurs reste depuis les premiers travaux en docimologie l'un des arguments forts pour faire la preuve du manque de validité des épreuves scolaires. Cependant, les docimologues, en abordant l'évaluation sous l'angle de la mesure et donc de la validité d'un score, n'ont traité finalement que de la validité de la notation. Depuis, les débats récurrents sur l'usage ou l'abandon des notes, outre les empoignades idéologiques suscitées quant aux modèles pédagogiques impliqués, renforcent cette focalisation excessive ou exclusive sur une petite partie des processus d'évaluation qui masque d'autres questions que nous jugeons bien plus importantes, si l'on entend produire des évaluations « au service des apprentissages »¹. Si les inquiétudes sur la fidélité des épreuves scolaires sont légitimes, elles ne constituent peut-être pas la question prioritaire. Celle-ci porterait plutôt sur la validité des évaluations, comme le remarquait bien avant nous Cardinet en 1973.

L'ambition principale de cet article est de montrer qu'une conjonction de travaux en didactiques et en psychologie permet de renouveler profondément les débats sur la validité. Les approches que nous qualifions d'« approches psycho-didactiques des évaluations » (APDE) bousculent, notamment, un certain nombre de catégories qui se sont progressivement imposées comme légitimes dans les travaux sur l'évaluation. L'entrée dans les questions d'évaluation par l'analyse de l'activité des évalués et des évaluateurs – caractéristique des APDE, avec l'importance accordée à la prise en compte effective des contenus² évalués – s'accompagne, entre autres, d'une prise de distance avec les typologies classiques des interventions d'évaluations (Scallan, 2000), en particulier, en mettant au second plan l'opposition entre sommatif et formatif³. Cette centration sur l'activité, c'est-à-dire à la fois sur ce qui est en jeu dans la réponse et le processus de réponse à une tâche d'évaluation, questionne également les discours sur les oppositions entre types d'épreuves.

Dans la première partie, après avoir résumé ce que l'on pourrait attendre d'épreuves scolaires valides, nous proposons un aperçu des travaux précurseurs qui, selon nous, s'inscrivent au sein des APDE. En référence à celles-ci, nous présentons dans la deuxième partie quelques principes qui permettent d'enrichir l'analyse de la validité des épreuves et qui fondent la « validité psycho-didactique » (VPD). La troisième partie est constituée d'illustrations et d'analyses, portant sur des tâches d'évaluation en mathématique et en lecture, qui témoignent de l'intérêt de la VPD. Dans la quatrième et dernière partie, nous présentons les deux axes qui permettent de définir et de s'assurer de la VPD des épreuves, puis nous abordons quelques pistes de réflexions nous permettant de revenir sur la raison qui motive ce texte et que nous venons de mentionner : s'assurer de la validité des épreuves.

2. La validité des épreuves scolaires : que peut-on attendre d'une évaluation scolaire ?

2.1. L'objet de l'évaluation et ses spécificités

La validité des épreuves scolaires est nécessairement pluridimensionnelle et doit être pensée à partir des visées d'une évaluation scolaire. Avec Bodin (1985, p. 17), nous posons comme principe « *que l'évaluation pédagogique se donne pour but la connaissance, avec le moins d'incertitude possible, du savoir de l'élève* ». En complément, nous ajoutons que les connaissances ou compétences en jeu doivent guider les choix d'évaluation en fonction de spécificités attachées à leur domaine. Evaluer en mathématiques, par exemple la maîtrise de la proportionnalité lors de la résolution de problèmes, n'est que très peu comparable à l'évaluation de la compréhension d'un texte. Les spécificités relatives aux disciplines et aux contenus impliqués ont une incidence

¹ Ce texte émane en partie d'une conférence donnée lors de la 9^{ème} journée de l'ABC-éduc (Association belge francophone des chercheurs en éducation) à Nivelles le 21 octobre 2013. Il s'agissait de réfléchir aux conditions d'une évaluation au service des apprentissages.

² Nous adoptons dans ce texte la définition des didacticiens : la notion de contenus « désigne tout ce qui est objet d'enseignement et d'apprentissage et qui constitue les savoirs qui sont enseignés et les connaissances que construisent les élèves au fil du temps » (Delcambre, 2007, p. 45).

³ Pour dire les choses autrement, nous rejoignons De Ketele (2006), au moins au niveau de sa démarche, lorsqu'il propose de distinguer pour caractériser l'évaluation la fonction principale visée (orientation, régulation, certification) et le type de démarche utilisée (sommative, descriptive, herméneutique). Par exemple, pour cet auteur, la démarche descriptive serait par un phénomène de réductionnisme historique restreinte à la fonction régulatrice ou formative (De Ketele 2006, p. 103).

marquée sur la nature de l'activité de réponse des évalués, sur la diversité des enjeux cognitifs et conceptuels. Ignorer cela conduirait à ne pas prendre en compte l'activité de l'élève qui est pourtant une clé essentielle si l'on souhaite développer des évaluations qui portent effectivement sur ce qu'elles mettent en jeu sur le plan des savoirs, connaissances, compétences, etc.

2.2. La portée diagnostique des épreuves scolaires

À cet aspect essentiel qu'est la prise en compte des contenus et de leurs spécificités, s'en ajoute un second tout aussi essentiel : la portée diagnostique de l'épreuve. Une épreuve scolaire au service des apprentissages doit, du mieux qu'elle le peut, permettre de comprendre l'activité de réponse de l'élève. Aussi précis soit-il, le seul constat de la maîtrise d'un savoir ou d'une compétence est insuffisant au plan diagnostique s'il laisse dans l'ombre le processus qui conduit à la réponse. Cependant, ceci ne signifie pas qu'il faille privilégier ou évincer certains types de tâches ; en effet, un QCM construit avec des distracteurs orientés par une finalité diagnostique peut autant satisfaire cette finalité qu'une tâche ouverte plus ou moins bien « corrigée ».

2.3. Validité de conception et de correction des épreuves

La validité ne saurait être confinée ni à un type de tâches, ni à la seule conception des épreuves. Le dispositif de correction quel qu'il soit joue également un rôle important en terme de validité des épreuves scolaires. La correction – qu'elle soit très ouverte comme lors de la correction de dissertations, ou très contrainte comme pour certains outils où les erreurs sont codées en vue de décrire un profil de réponses dans une perspective diagnostique – devrait être pensée en ayant comme finalité principale la qualification de ce qui est réalisé par l'élève. L'évaluation est dans cette perspective une qualification au sens d'une « valeur ajoutée », celle qui consiste à donner du sens à ce qui est produit par l'élève ; autrement dit, elle n'est pas une disqualification de l'élève. Nous suivons ici Cardinet (1991) qui affirmait à propos de l'évaluation formative qu'elle devait dire quelque chose à l'élève et pas quelque chose de l'élève. Il invitait à « évaluer sans juger » (1989), nous inviterions, pour notre part, à « évaluer pour qualifier ». Une évaluation n'apportant qu'un constat d'échec ou de réussite sans une intégration dans une évolution propre à chaque apprentissage, même si elle peut être parfois légitime, est à un niveau individuel d'une grande pauvreté descriptive⁴.

2.4. Evaluation et modèles de développement

Toute évaluation repose (ou devrait reposer), plus ou moins explicitement, au minimum sur une hiérarchie de difficultés ou mieux encore sur un modèle de développement lorsqu'il existe. Cette distinction est fondamentale car dans le premier cas, on peut se contenter de déterminer, sans comprendre ce qui est en jeu, des niveaux de réussite. Dans le second cas, on doit, pour décrire les évolutions, se référer aux travaux, théories, modèles spécifiques à chaque apprentissage. Si la mise en relation de collections d'échecs et de réussites a permis à la psychométrie de proposer des hypothèses plus ou moins stables sur les structures de l'intelligence, dans le cas des apprentissages scolaires, il serait toutefois regrettable d'ignorer l'apport des travaux en didactiques et en psychologie cognitive. L'approche corrélationnelle peut du reste être trompeuse au plan diagnostique car des corrélations agrègent parfois des procédures ou processus de nature très différente. Ainsi une échelle unidimensionnelle au sens où on l'entend en psychométrie peut cacher une forme de développement en escalier. La réduction à une dimension et la construction d'une métrique, quel que soit le procédé statistique, tend à « linéariser » des évolutions qui sont pourtant qualitativement très différenciées.

2.5. L'intérêt de l'activité de réponse des évalués

Enfin, un critère fondamental pour la validité des épreuves scolaires est l'intérêt de l'activité suscitée par les tâches d'évaluation. De ce principe, rien de manichéen ne peut être tiré en termes de types de tâches. Il serait réducteur de confiner les débats sur la validité à la plus ou moins grande ouverture ou complexité d'une tâche. Une tâche de sélection de réponse peut susciter des élaborations cognitives plus importantes que certaines réponses écrites apparemment plus ouvertes. L'intérêt des tâches doit être pensé de manière plurielle, en fonction

⁴ Là encore, il ne faut pas perdre de vue la finalité dévolue aux outils : certaines tâches, que nous rejeterions d'un point de vue psycho-didactique, peuvent s'avérer efficaces (pertinentes) s'il s'agit de savoir si le niveau baisse ou s'il monte, tout en ne présentant absolument aucun intérêt pour ce qui concerne la régulation des apprentissages. On retrouve la distinction opérée dès 1973 par Cardinet entre tests à visée collective et tests à visée individuelle. Pour les premiers, il s'agit de s'occuper de l'échantillonnage des élèves, le choix des questions ne pose pas de problème, une question unique peut très bien donner lieu à une enquête.

au moins de l'intérêt diagnostique (cf. paragraphe précédent), mais aussi de l'intérêt formatif. Renoncer à la dissertation au motif qu'elle provoque des divergences de notation serait appauvrissant pour les élèves qui ont besoin d'être confrontés à ce type de tâches, notamment à travers des situations d'évaluation, afin de développer et d'enrichir leurs capacités dans le domaine de la production d'écrit et de l'argumentation (Goasdoué, Vantourout et Bedoin, à paraître).

3. Pour des approches psycho-didactiques des évaluations (APDE)

Les travaux sur l'évaluation formative ont de longue date mis en avant l'intérêt « *de comprendre avant tout le fonctionnement cognitif de l'élève face à la tâche proposée* » (Allal, 1979, p. 136). Cet objectif qui caractérise les APDE ne peut être atteint en ignorant les apports conjoints des didactiques disciplinaires et de la psychologie cognitive. Les travaux en didactiques permettent de cerner les enjeux conceptuels des tâches ; la description des procédures et conceptions⁵ des élèves dans un champ de savoir donné qu'ils proposent est également essentielle pour comprendre le fonctionnement cognitif de l'élève. Les didacticiens à travers l'idée et la réalisation d'analyses a priori ont également enrichi l'analyse/formalisation des tâches issue de travaux en psychologie, en particulier ceux sur la résolution de problèmes. Les travaux en psychologie cognitive et du développement offrent par ailleurs des concepts essentiels, tels que la notion de schème, de représentation, très utiles pour analyser l'activité des élèves évalués. Faute de place, nous ne pourrions décrire l'histoire de ces convergences fructueuses et parfois conflictuelles entre psychologie et didactique. Retenons cependant quelques caractéristiques essentielles des approches psycho-didactiques des apprentissages (Weil-Barais 1996 ; Maury 2001 ; Maury 2005) susceptibles de renouveler le regard sur les tâches d'évaluation⁶. Ces approches, d'une part, sont structurées par la question de l'enseignement-apprentissage de connaissances identifiées dans un champ défini, et, d'autre part, reposent sur une « *une conception développementale du processus d'enseignement-apprentissage, qui s'appuie sur une analyse des processus psychologiques, en regard des caractéristiques épistémologiques des savoirs en jeu* » (Goffard et Weil-Barais, 2005, p. 133). Transposé à l'évaluation ces deux traits fondamentaux conduisent d'une part à ne pas chercher un trait latent, ou une compétence mal définie⁷, mais à décrire pour un champ donné la maîtrise de savoirs ou savoir-faire et d'autre part à considérer les résultats de l'évaluation non pas comme une description quantitative de l'efficacité d'une compétence, mais plutôt comme la marque d'un niveau de développement (ou d'acquisition d'un concept, par exemple). Cette conception remet en partie en cause l'idée d'échelle au sens d'une métrique, car le développement est le plus souvent caractérisé par des réorganisations qualitatives importantes et rien n'indique que le pas entre un degré et un autre dans un score est homogène. Brièvement, les APDE sont proches de la manière ambitieuse dont Vergnaud (2001) conçoit l'évaluation des compétences. Pour cet auteur, « *l'évaluation des compétences appelle non seulement l'analyse du résultat de l'activité, mais aussi celle de l'organisation de l'activité* » (Vergnaud 2001, p. 46). Il propose une méthode d'analyse qu'il qualifie de « *"descente vers le cognitif" : pas d'analyse des compétences sans analyse de l'activité, pas d'analyse de l'activité sans analyse des conceptualisations sous-jacentes* » (Vergnaud 2001, p. 47). L'objectif de « *descente vers le cognitif* » engage à aller au-delà de la performance et à s'intéresser aux processus de réponse des évalués.

3.1. Travaux s'inscrivant au sein des APDE

Nous distinguons au sein de l'activité des évaluateurs deux facettes, la conception d'épreuves et leur correction au sens large. Bien qu'ils soient encore peu nombreux, des didacticiens des mathématiques abordent l'évaluation avec des principes voisins de ceux que nous regroupons sous l'étiquette « APDE ». Dans leurs travaux, ils s'intéressent principalement, à notre connaissance, à la conception d'épreuves. Concernant la

⁵ La notion de conception est souvent associée à celle de procédure. Elle est utilisée pour parler du rapport des sujets à des concepts. Les conceptions peuvent être définies en première approche comme des manières de connaître. Charnay (1996, p. 39) en donne une définition synthétique: « *La notion de conception [rend] compte de la façon dont l'élève s'est approprié une connaissance donnée et de la manière (correcte ou non) qu'il a de la comprendre* ». Cette notion « *aide à mieux élucider le fonctionnement et les difficultés des élèves* » (Charnay, 1996, p. 39). Il n'existe pas de définition générale de la notion de procédure, il faut souvent la reconstruire en fonction de ce qui est observé, de la situation et des contenus impliqués.

⁶ Les APDE sont une adaptation pour le champ de l'évaluation des « *approches psycho-didactiques des apprentissages* » (APDA).

⁷ Ce terme n'est pas péjoratif, il est utilisé dans le sens qu'il a en résolution de problème : un problème mal défini est un problème aux contours flous.

correction d'épreuves, nous mentionnerons quelques-uns de nos travaux qui portent sur les mathématiques et les sciences économiques et sociales (SES).

3.1.1. *La conception d'épreuves*

A la suite de Bodin (1997, 2005) et des études organisées dans le cadre de l'observatoire EVAPM⁸, Grugeon-Allys et son équipe (Grugeon-Allys, Pilet, Chenevotot-Quentin et Delozanne, 2012 ; Chenevotot-Quentin, Grugeon-Allys, Pilet et Delozanne, 2012) ont développé au cours de plusieurs recherches finalisées, un outil d'évaluation diagnostique⁹, dont l'originalité et la force sont de s'appuyer sur un modèle multidimensionnel de la compétence algébrique en fin de scolarité obligatoire (16 ans, collège ou lycée). Le diagnostic développé vise – effectivement (c'est nous qui soulignons) – une analyse globale multidimensionnelle des connaissances et de l'activité des élèves en algèbre (Grugeon-Allys et al., 2012). « *Il s'appuie sur une évaluation des réponses des élèves (erreurs et procédures) et un diagnostic cognitif global [... fondé] sur une étude cognitive et épistémologique de l'algèbre élémentaire élargie à des études empiriques* » (Grugeon-Allys et al., 2012, p. 5). Le point de départ de ces travaux a été de définir le champ conceptuel de l'algèbre élémentaire en fin de scolarité obligatoire (Grugeon, 1997). En référence à la théorie anthropologique du didactique de Chevallard, « *chaque tâche diagnostique est caractérisée par : un type de tâche, les registres sémiotiques et la complexité des objets algébriques mis en jeu, le niveau d'intervention des organisations mathématiques dans la tâche, des techniques attendues relativement aux éléments technologiques et théoriques visés* » (Grugeon-Allys et al. 2012, p. 7). Bien qu'elle ne soit pas formulée en ces termes, la question de la validité est importante pour ces auteurs, avec une préoccupation particulière pour la validité de contenu (au sens de la psychométrie, c'est-à-dire de couverture de domaine). Cependant ce terme n'a pas exactement le même sens qu'en psychométrie car cette validité est pensée a priori à partir d'une analyse épistémologique d'un champ de savoir et non a posteriori par confrontation avec d'autres épreuves.

3.1.2. *La correction d'épreuves*

L'intérêt que nous portons à la validité est lié au départ à une série de travaux sur l'activité de professeurs confrontés à l'évaluation de tâches complexes en mathématiques (Vantourout, 2007 ; Vantourout 2004 ; Vantourout et Maury, 2006 ; Nabbout, 2006). De ces travaux, il ressort un résultat que l'on peut résumer ainsi : la diversité des jugements évaluatifs ne peut être comprise qu'à partir du moment où l'on s'intéresse aux connaissances que les évaluateurs mobilisent, ou pas, lors de leur activité. Un même jugement évaluatif peut reposer sur des connaissances différentes, et des connaissances analogues peuvent produire une diversité de jugements (Vantourout, 2004). Ce résultat a été confirmé pour la correction de dissertations en SES (Vantourout et Goasdoué, 2010). Nous le répétons, la question de la fidélité ne nous paraît pas devoir être première : c'est un fait, il existe un décalage entre les jugements et les notes produites car ces dernières ne pourront jamais résumer la richesse des premiers. Pour ce qui est de l'activité de correction, les APDE privilégient la validité au détriment de la fidélité. À un niveau très général, nous avons proposé de considérer la validité de la correction comme la qualité des inférences produites à partir d'éléments considérés comme des indices dans la copie. Ces inférences étant elles-mêmes dépendantes des connaissances du correcteur et de la régulation de son activité. Cette proposition rejoint assez nettement celle de la validité d'un diagnostic médical, dont l'une des qualités est de guider vers des explications des symptômes. Ceci suppose donc que les enseignants disposent d'un « modèle » du fonctionnement cognitif des élèves dans un domaine précis. Enfin, la validité de la correction doit être mise en lien avec les objectifs de l'épreuve et les finalités que s'assigne le correcteur au fil de sa correction (Goasdoué, et al., à paraître).

3.1.3. *Des tests psychométriques et didactiques à visée diagnostique*

Soucieux d'améliorer la portée diagnostique des tests, des psychométriciens font appel à la didactique d'une part pour analyser la difficulté des tâches (Riopel, Sakr, Raïche, Potvin et Léocadie Djédjé, 2011) et d'autre part pour affiner l'analyse des processus cognitifs en jeu dans un test, dans le cadre de préoccupations relatives à sa validité diagnostique (Loye, Caron, Pineault, Tessier-Ballairgeon, Burney-Vincent et Gagnon, 2011). Ces travaux, dont les apports respectifs sont manifestes, invitent cependant à des précisions concernant ce que

⁸ EVAPM est une recherche conduite depuis le milieu des années quatre-vingts qui porte sur les acquis des élèves de l'enseignement secondaire français en mathématiques (voir Bodin, 1997, 2005).

⁹ Ce outil d'évaluation diagnostique est désormais un test informatisé, disponible sur la plateforme MathEnPoche de l'association Sésamath : <http://mathenpoche.sesamath.net/>

recouvre l'expression « processus cognitifs ». Chez des didacticiens (Grugeon-Allys et al., 2012) et, dans une moindre mesure, chez des personnes qui leur sont proches (Demonty, Fagnant et Dupont, soumis), cette expression renvoie à des processus spécifiques identifiés au cours d'une analyse impliquant des contenus précis en relation avec des types de tâches. Pour d'autres auteurs, ceux inscrits dans le champ de la psychométrie, elle renvoie à des processus bien plus généraux. Loye et son équipe (2011, p. 16 et 27), pour réaliser une classification des items qu'ils qualifient de « didactique », s'appuient sur une grille qui présente les quatre phases du processus de résolution de problème (analyse, planification, exécution, retour)¹⁰. De leur côté, Riopel et al., (2011, p. 57-59), qui s'intéressent aux équations du premier degré à une ou deux inconnues, parviennent à dégager neuf variables qui permettent de prédire le niveau de difficulté de tâches¹¹, en s'appuyant pour certaines d'entre elles sur des travaux « reconnus » en didactique des mathématiques, entre autres ceux consacrés aux différents modes de représentation des fonctions. Mais là aussi, ce qui relève des processus cognitifs semble très général et renvoie à la définition (sans aucune précision) de trois niveaux d'habileté cognitive : habileté d'application, conceptuelle, de résolution (Riopel et al., 2011, p. 66). Le même reproche nous semble pouvoir être adressé à Bodin (2005, p. 5-6) qui analyse les demandes cognitives de diverses questions d'évaluation (dont celles de PISA 2003) en se basant sur une taxonomie dont les catégories principales, proches des précédentes, sont : A) connaissance et reconnaissance, B) compréhension, C) Application, D) Créativité. On notera que cette dernière taxonomie s'inspire largement, de celle de Bloom, auteur qui ne s'est jamais engagé dans une analyse de processus cognitifs effectifs. Elle est proche de celle utilisée par des spécialistes de la mesure qui font, en accord avec leurs priorités, l'impasse sur les contenus (Laveault et Grégoire, 2002, p. 20-21). Tout se passe comme si les analyses a posteriori, qui portent sur des tâches que les chercheurs n'ont pas conçues mais qu'ils « récupèrent », ne pouvaient que déboucher sur des analyses très globales des processus cognitifs, par opposition aux travaux qui reposent sur l'élaboration de tâches et qui de ce fait peuvent faire des hypothèses plus précises sur le fonctionnement cognitif notamment à travers l'analyse a priori des tâches. On retrouve ici une opposition de fond entre psychologie cognitive et différentielle : les tâches sont conçues soit comme des moyens de mettre en avant le fonctionnement cognitif (conception expérimentale des tâches), soit c'est la réussite à plusieurs tâches qui permet d'inférer l'existence d'un trait latent (conception différentielle). Si l'approche psychométrique peut avoir un intérêt pour décrire les structures et processus latents pas encore clairement définis comme ceux caractérisant l'intelligence, dans le champ de l'évaluation scolaire et plus encore quand il s'agit d'un domaine précis, ce choix est nettement plus discutable. Pourquoi continuer d'ignorer, pour évaluer les acquis des élèves, les descriptions que des didacticiens et psychologues font des procédures ? À nouveau, parce que le but n'est pas d'évaluer un trait latent propre à un élève mais ses acquis, ces travaux constituent un apport considérable à la définition de ce qui est « mesuré » et à sa signification. Deux questions fondamentales concernant la validité – « que mesure cette épreuve ? » ; « que puis-je faire de cette mesure ? » (De Landsheere, 1988, p. 75) – auxquelles le développement récent de la psychométrie dans l'évaluation scolaire, via les modèles de réponse à l'item, est en permanence confronté (Vrignaud, 2006).

Pour résumer notre position, nous estimons que la conception d'épreuves scolaires, ainsi que leurs analyses et leur portée diagnostique, peuvent être notablement enrichies en dépassant la simple prise en compte des performances en termes de réussites et d'échecs au profit d'une analyse de l'activité des élèves sous-jacente à ces résultats.

3.2. Validités et validité psycho-didactique¹² (VPD)

La définition de la validité est souvent résumée par une question : « l'épreuve mesure-t-elle bien ce qu'elle est censée mesurer ? ». Trois remarques s'imposent à propos de cette interrogation apparemment univoque et

¹⁰ Cette catégorisation renvoie à Polya (1945, *How to solve it ?*), mathématicien dont les apports « didactiques » sont remis en cause par les didacticiens des mathématiques et par Julo (1995, p. 65) : « *En fait [la manière de Polya] d'analyser la démarche de résolution est celle du mathématicien qui sait résoudre le problème et ne rend aucunement compte de ce qui se passe au niveau des processus cognitifs* ».

¹¹ Parmi ces variables figurent les caractéristiques des données du problème (présence ou non de l'ordonnée à l'origine) et des paramètres de l'équation (signe de la pente), l'habileté cognitive visée (conceptualisation, application et résolution), le type de réponse (à choix multiples, courte, élaborée) (Riopel et al, 2011, p. 68-69).

¹² Ce que nous qualifions de VPD ne décrit pas prioritairement, bien qu'elle puisse y contribuer, l'intérêt qu'une évaluation peut représenter pour l'enseignement. La VPD doit ainsi être distingué du sens attribué par Bain (2003) à l'expression validité didactique. Bien que les usages didactiques (au sens plutôt d'intérêt et d'impact pour l'enseignement, si l'on reprend l'idée de Bain) d'une évaluation dépendent nettement de sa validité psycho-didactique, la focalisation sur l'immédiat de l'activité de réponse ne permet pas de traiter simultanément des exploitations ultérieures possibles.

pleine de bon sens. D'abord, le terme « ce » laisse très clairement apparaître l'indépendance entre la mesure et l'objet mesuré. Cette formule régulièrement employée dans les manuels de psychométrie est pour le moins contradictoire avec une autre tout aussi ambiguë : « l'intelligence, c'est ce que mesure mon test ». Le caractère autoréférentiel de l'idée même de construit est en effet opposée à cette présumée indépendance de la mesure et de son objet, d'autant plus que la validité de contenu est abordée de manière inductive en psychométrie, par des analyses corrélationnelles. Ensuite, cette formulation fait fi du processus qui conduit au résultat. C'est considérer que ce qui est mesuré, non seulement existe, mais également qu'il n'est pas influencé par la manière dont on le mesure. Si l'interaction entre le mètre qui sert à mesurer et ce qui est mesuré est négligeable, ce n'est pas le cas de l'interaction de l'élève avec les situations qui lui sont proposées à des fins d'évaluation. Sans aller jusqu'à transposer dans le champ des évaluations scolaires ce que Bourdieu affirmait à propos des sondages, « l'opinion publique n'existe pas », il faut cependant rappeler que toute évaluation construit ce qui est observé. D'où l'importance de maîtriser ce qui est en jeu dans les évaluations. Enfin, la formulation reprise en début de paragraphe masque une autre définition implicite de la validité qui pourrait être formulée ainsi : « le score décrit-il bien le trait latent qui est censé sous-tendre les performances observées ? ».

3.2.1. Principe 1 : définition a priori de ce qui est mesuré

Le premier principe qui fonde la validité psycho-didactique est que l'évaluation n'a pas pour but de décrire le niveau d'un individu sur une échelle décrivant à son tour une position sur un trait latent, mais de décrire/qualifier le fonctionnement cognitif à partir de l'activité de réponse qui est pour une par déterminée par le couple sujet-situation. Ainsi, maîtriser l'algèbre n'est pas parvenir à un certain niveau sur un trait latent, mais c'est le résultat de la coordination de multiples concepts et procédures qui agrégés peuvent être résumés par le terme « compétence algébrique » (Grugeon, 1997). L'analyse de la validité ne peut donc se passer d'une analyse a priori des savoirs, connaissances, concepts et processus à l'œuvre dans un apprentissage. La VPD se démarque ainsi nettement d'autres procédés de validation qui s'intéressent prioritairement au score et reposent sur des analyses « statistiques » de la validité. Enfin, la VPD en focalisant son attention sur l'interaction sujet-tâche ne traite pas de la « couverture du domaine ». Grugeon-Allys et al. (2012) ont donné un sens nouveau à ce qui pourrait qualifier de validité de contenu, à l'occasion de l'élaboration de leur test PEPITE (voir supra 3.1.1). Les items élaborés représentent un « échantillon représentatif » mais qualitativement défini des tâches utilisées dans l'enseignement et caractérisent la diversité des aspects de la compétence algébrique. Cette forme qualitative de la validité de contenu, également fondée sur des travaux en didactique, est nettement complémentaire à la VPD et pourrait être qualifiée de « validité épistémologique ».

3.2.2. Principe 2 : spécificité / généralité

L'opposition classique « spécificité/généralité » est une question qui traverse les APDE et que la VPD cristallise. Située à la croisée du didactique et du méthodologique, lesquels renvoient respectivement au spécifique et au général, la VPD témoigne de la manière dont il est possible de dépasser cette opposition en les articulant. La démarche d'analyse permettant de s'assurer de la VPD, parce qu'elle repose sur l'analyse de l'activité, est applicable à toute évaluation, indépendamment de sa fonction et de ses modalités. Toutefois, la variété des activités de réponse suscitées par chaque tâche nécessite des déclinaisons spécifiques et des références théoriques différentes. Nous considérons que toute démarche méthodologique, qui comporte, pour s'assurer de la validité d'une épreuve, une analyse fondée sur des travaux en psychologie cognitive et/ou en didactique disciplinaire en lien avec les contenus impliqués dans les épreuves, peut être qualifiée de psycho-didactique, et ce quel que soit le domaine ou la discipline évalué. Par ailleurs l'orientation ou la profondeur (en référence à la « descente vers le cognitif ») des analyses conduites est spécifique des fonctions assignées aux tâches d'évaluation. Ainsi, certaines tâches ou formulations permettent par exemple de mieux cerner les procédures de résolution adoptées en mathématiques ou le niveau de la représentation élaborée lors de la compréhension d'un texte, ce qui est fondamental dans une visée diagnostique. Dans une finalité certificative, on peut être amené à chercher, au contraire, la ou les tâches aillant la plus grande portée de « généralisabilité qualitative » des résultats. Pour le dire sommairement, il existe des tâches, par exemple en algèbre, pour lesquelles on sait qu'elles ne peuvent être réussies que par des élèves qui maîtrisent les divers aspects de la compétence algébrique.

3.2.3. Principe 3 : une entrée par l'activité

Comme nous l'avons mentionné à plusieurs reprises, l'étude de la VPD repose sur l'analyse de l'activité de réponse, « imaginée » ou effective, à des tâches d'évaluation. Ce choix permet, comme ce fut le cas historiquement en psychologie d'échapper au schéma stimulus – réponse. L'activité de réponse dans n'importe quelle forme d'évaluation est un tout structuré qui repose sur la coordination de différentes ressources qu'on ne peut que très difficilement étudier isolément. Les procédés usuels d'analyse des items en psychométrie sont, à ce sujet, intrigants puisque les scores ne sont pas seulement la quantification d'une réponse à une stimulation mais

un moyen de déterminer ce qui caractérise le stimulus. La VPD se rapproche ainsi nettement plus des conceptions expérimentales en psychologie où les stimuli sont construits dans une finalité précise, celle de décrire indirectement les processus cognitifs en jeu. Cependant, à la différence de tâches expérimentales qui ne visent qu'un aspect très restreint du fonctionnement cognitif, toute épreuve scolaire est à la fois beaucoup plus riche et aussi en un sens moins précise, ce qui nécessite une analyse fine de l'activité.

4. Quelques aspects de la validité psycho-didactique : des exemples pour cerner les enjeux

Nous considérons qu'il n'existe pas de tâches valides en soi mais des regards différents sur la validité relativement à des finalités d'évaluation. La finalité principale de la VPD est avant tout d'assurer la qualité du diagnostic, ce qui peut conduire à se poser les questions suivantes :

- la tâche implique-t-elle bien la mise en jeu des connaissances que l'on cherche à évaluer ?
- la tâche ne sollicite-t-elle pas de « compétence écran » ?
- Le choix des valeurs affectées aux variables didactiques influence-t-il les choix de réponses, les procédures ? va-t-il jusqu'à transformer profondément la tâche et l'activité de réponse ?
- Certaines réponses ne sont-elles pas privilégiées du fait des modalités de présentations du dispositif de réponse ? par le choix des distracteurs opérés par les concepteurs ? par les relations entre différentes questions ?
- Quelle est la distance entre une tâche d'évaluation et les habitudes d'enseignement relatives à ce qui est évalué ?
- L'activité de réponse peut-elle conduire à un remaniement de ce qui est connu, appris ou lu ?

Nous allons présenter quelques illustrations relatives à la plupart de ces questions. Des illustrations peuvent renvoyer à divers degrés à plusieurs questions. En effet, si ces questions précédentes sont plutôt bien circonscrites, il s'avère que, dès que l'on s'engage dans l'analyse effective d'une épreuve, elles concernent assez souvent des caractéristiques qu'il est difficile d'aborder de manière totalement séparée. Chaque sous-titre renvoie donc à la question que nous voulons prioritairement illustrer à partir de la tâche analysée. L'attention portée à l'analyse des tâches étant la « marque de fabrique » de la VPD, nous développerons certaines analyses afin de rester cohérents avec ce que nous défendons.

4.1.1. La tâche met-elle bien en jeu les connaissances évaluées ?

Mercier (2007, p. 95-96), en analysant un item extrait de PISA où il s'agit, selon les concepteurs, d'une application du théorème de Thalès, montre que des élèves peuvent répondre à une situation et réussir massivement sans mobiliser la connaissance attendue. Bodin (2006) critique lui aussi certaines attentes de PISA relativement à la prise en compte des connaissances mobilisées. Il cible la deuxième question du problème « Pommiers » correspondant à la situation d'un jardin planté de pommiers et de conifères. Nous nous appuyons ci-après sur sa présentation (Bodin, 2006, p. 12) ; l'item dont il est question est le suivant :

Question 2 : POMMIERS

M136Q02-00 11 12 13 14 15 99

Il existe deux expressions que vous pouvez utiliser pour calculer le nombre de pommiers et le nombre de conifères dans cette situation :

Nombre de pommiers = n^2

Nombre de conifères = $8n$

où n est le nombre de rangées de pommiers.

Il existe une valeur de n pour laquelle le nombre de pommiers est égal au nombre de conifères. Trouvez cette valeur de n et expliquez votre méthode pour la calculer.

Figure 1. Énoncé extrait de l'enquête PISA 2000 – Exercices libres de diffusion en culture mathématique

La mathématisation est quasi évidente et conduit à une équation à résoudre dans \mathbb{N} : $n^2 = 8n$. Une partie des élèves fait une résolution mathématiquement correcte, reposant sur une factorisation : $n^2 = 8n$; $n^2 - 8n = 0$; $n(n - 8) = 0$; $n = 0$ ou $n = 8$. Ils trouvent deux valeurs 0 et 8, puis éliminant la valeur 0, ils reconnaissent que la valeur 8 (qui signifie qu'il y a 64 pommiers et 64 conifères) est la seule qui convient. Une autre partie des élèves donnera une réponse correcte en utilisant cette fois une démarche de résolution incorrecte « en simplifiant par n » : $n^2 = 8n$; $n \times n = 8 \times n$; $n \times \cancel{n} = 8 \times \cancel{n}$; $n = 8$. Ces deux procédures ont été considérées comme correctes,

alors que l'une d'elle est incorrecte mathématiquement (car elle revient à diviser par n sans supposer n différent de 0, après avoir remarqué que $n=0$ est une solution de l'équation dans \mathbb{N} , mais pas vraiment une solution du problème qui suppose qu'il y a au moins un pommier). Par ailleurs, certains peuvent utiliser une procédure arithmétique qui consiste à tester l'ensemble de valeurs de n (n étant petit, cette procédure est peu coûteuse). Bodin (2006, p. 12) soulève à juste titre la nécessité « *d'études qualitatives complémentaires destinées à analyser d'un point de vue mathématique et didactique les procédures suivies par les élèves* ». Nous reformulerons aussi cela en posant une question : pour certains élèves ayant « réussi », qu'évalue effectivement cette évaluation ?

4.1.2. Choix des valeurs affectées aux variables didactiques

Nous emprunterons cet exemple à Levain (1997) dont les travaux sur le développement cognitif de la proportionnalité relèvent sans conteste des approches psycho-didactiques des apprentissages. Cet auteur s'est intéressé aux problèmes d'agrandissement/réduction et d'échelle. Il s'agit de problèmes complexes au sens où ils peuvent mobiliser des cadres géométrique, arithmétique, numérique et algébrique (Levain 1997, p. 92). Levain propose un premier problème (Levain 1997, p. 108) :

Imagine une chambre idéale et donne ses dimensions :

Cette chambre mesure mètres de long et mètres de large.

Calcule une échelle qui te permette de représenter cette chambre

Figure 2. Énoncé du problème n°13 – Levain, (1997), p. 108

Ce problème est extrêmement complexe (voir l'analyse de Levain en termes de procédures, 1997, p. 202-216). Il nous intéresse ici car il peut être résolu de manière « astucieuse » (Levain 1997, p. 201). En effet, un élève peut répondre que sa chambre idéale mesure 2,1 m x 2,97 m ou 21 m x 29,7 m. Dès lors, la plus grande part de difficulté du problème disparaît et les connaissances mobilisées ne sont plus du tout de même nature.

Levain (1997, p. 201) propose une seconde version du problème en imposant des dimensions afin d'empêcher toute solution astucieuse (il modifie la valeur d'une variable didactique) :

Tu veux faire le plan d'une salle des fêtes qui mesure

58 mètres de long et 28 mètres de large.

Calcule une échelle pour que ton plan tienne tout entier

dans une feuille comme celle-ci (format 21 x

Figure 3. Énoncé du problème n°13bis – Levain, (1997), p. 201

Nous considérons que les conséquences de cette modification, au niveau du fonctionnement cognitif, témoignent de l'intérêt d'étudier la validité d'un exercice d'un point de vue psycho-didactique : quels sont les connaissances et les processus de réponse qui vont être activés ? Ce problème dans ces deux versions successives permettra d'évaluer « la même chose » pour des élèves « non astucieux », mais pour les élèves astucieux il en ira tout autrement.

4.1.3. Les modalités de réponses

Dans la thèse de Blanc (en cours), une partie du corpus rassemble cinquante exercices utilisés par dix-neuf enseignants de CP (cours préparatoire, première année de l'école élémentaire) pour évaluer la compétence¹³ « Connaître les correspondances entre les lettres et les sons ». Les résultats obtenus à ces exercices ont servi à renseigner les livrets scolaires des élèves des classes où ils ont été recueillis. Cette compétence,

¹³ Compétence ou item ; ces termes sont synonymes dans le cadre du « Livret scolaire de l'école primaire » où figure cette compétence. En France, à la fin de chaque trimestre, les enseignants du primaire doivent remettre ce livret aux parents.

comparativement à beaucoup d'autres, est très largement évaluée, d'une part, parce que l'enjeu est important au regard des connaissances actuelles et consensuelles sur l'apprentissage de la lecture, d'autre part et certainement, parce qu'elle est (ou semble) aisée à identifier et à circonscrire. Concernant l'évaluation des élèves, l'enjeu est également majeur : une évaluation valide doit permettre à l'enseignant d'identifier parmi ses élèves ceux qui ont acquis et maîtrisent cette compétence mais surtout ceux pour lesquels ce n'est pas encore le cas. L'objectif est d'évaluer les connaissances des élèves en matière de codage phonologique au moyen de graphèmes¹⁴ : quel phonème (sortie) code le graphème donné (entrée) ? Evaluer cette compétence paraît a priori relativement simple. En effet, il suffit pour cela de demander à des élèves de lire à voix haute une série de syllabes adéquatement choisies. En procédant ainsi, l'évaluateur met de son côté toutes les chances pour proposer une tâche valide. Il s'agit en quelque sorte de la tâche « fondamentale »¹⁵ pour évaluer cette compétence. Or, dans le corpus exhaustif, aucun exercice correspondant à une telle tâche n'a été trouvé. Mais, résultat beaucoup plus embarrassant, seulement six des cinquante exercices analysés par Blanc sont valides.

Ainsi l'un des enseignants propose un exercice qui pourrait renvoyer à ce qu'il déclare vouloir évaluer. Il suffirait d'utiliser le support, les cinq mots, en demandant aux élèves de les lire. Toutefois l'exercice proposé (figure 4) revient en réalité à un simple exercice de reconnaissance visuelle : il suffit simplement d'entourer toutes les lettres « i » pour réussir.

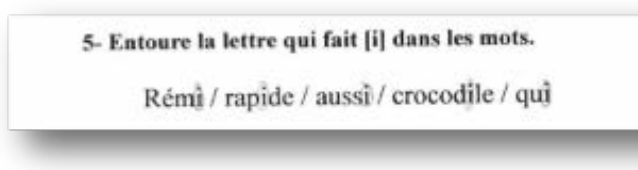


Figure 4. Exercice figurant dans le corpus de la thèse de C. Blanc

Prenons quelques autres exercices extraits du corpus.

Un premier enseignant propose (exercice A) :

2. Connaître les correspondances entre les lettres et les sons
Je numérote la lettre qui produit le son produit par le maître.

Compétence 10 : 8 / 10

i	u	p	r	a	l	e	s	m	n	o	t	j
		5	1	2			3					

Figure 5. Exercice figurant dans le corpus de la thèse de C. Blanc

¹⁴ On cherche ici à évaluer l'apprentissage explicite des correspondances graphèmes-phonèmes qui conduit les enfants à développer des relations entre les représentations des lettres et unités phonologiques larges telles les syllabes ou les infrasyllabes (attaque et rime) (voir Ecalle et Magnan, 2002, p. 42).

¹⁵ A l'instar de Brousseau (1998), pour qui « chaque connaissance peut se caractériser par une (ou des) situation adidactique qui en préserve le sens et [qu'il appelle] situation fondamentale » (Brousseau 1998, p. 60), nous pensons qu'il pourrait exister des tâches d'évaluation « fondamentales », c'est-à-dire valides d'un point de vue psycho-didactique au sens où le sujet doit absolument mobiliser la connaissance visée pour réussir. Afin d'aider les enseignants, il serait important de déterminer quelles sont ces tâches pour répondre à la commande institutionnelle en matière d'évaluation, avec le doute qu'il n'en existe pas pour couvrir l'intégralité de cette commande, ne serait-ce qu'au niveau de la scolarité obligatoire.

Un deuxième enseignant propose (exercices B et C) :

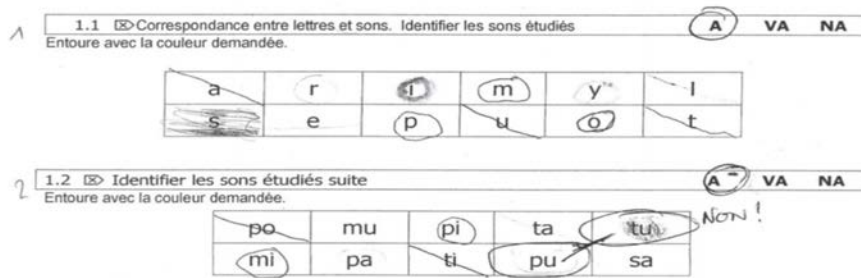


Figure 6. Exercices figurant dans le corpus de la thèse de C. Blanc

Beaucoup d'exercices, comme les exercices A, B et C, prennent la forme de « dictées aménagées » dans lesquelles l'élève n'a pas à traduire en oralisant le son proposé (phonème) mais à le retrouver en le reconnaissant au sein d'une liste fermée de propositions écrites (lettres ou syllabes), puis à laisser une trace écrite afin de permettre à l'enseignant d'évaluer en différé la réponse. Soulignons qu'il s'agit ici de reconnaître les correspondances entre les sons et les lettres, l'inverse de ce qui est annoncé et reporté dans les livrets. On comprend l'intérêt de ce type d'exercices pour l'enseignant car ils permettent une passation collective avec le même support pour tous les élèves. Dans ces exercices, comme dans la tâche fondamentale, pour réussir les élèves doivent lire correctement. Mais les activités qui accompagnent l'activité de lecture-décodage sont différentes, principalement parce que dans la tâche canonique le stimulus est visuel alors que dans les dictées aménagées il est sonore. Ces trois exercices se distinguent par la nature des unités à reconnaître, lettres ou syllabes. Certains choix de l'enseignant renvoient à des confusions ou difficultés classiques, chez des élèves de CP, dans le domaine des correspondances entre lettres et sons, qu'il s'agisse de confusions auditives ou visuelles. Mais, pour nous, présentement, ces exercices se distinguent également par leurs modalités de présentation, ici des tableaux plus ou moins étendus, ainsi que par leurs modalités de réponse, autrement dit par la nature de la trace écrite à produire. Les modalités de réponses doivent être considérées comme génératrices d'une tâche supplémentaire qui, même si elle paraît minimale, peut mettre en difficulté certains élèves, en particulier ceux en difficulté. A cette tâche supplémentaire sont associées des consignes multiples qu'il est nécessaire de comprendre et de mémoriser. Prenons l'exercice A : il est probable que l'enseignant accompagne ses élèves, qu'il leur demande d'inscrire le chiffre « 1 » lorsqu'il émet le premier son, il se peut même que ce chiffre soit écrit au tableau. Bien sûr, tous ces éléments renvoient à des apprentissages nécessaires pour réussir à l'école, mais nous voulons attirer l'attention sur le coût cognitif de ces activités supplémentaires. Celles-ci requièrent parfois la mobilisation de compétences qui ne sont pas celles prétendument évaluées, on parle alors de « compétences écran », au sens où leur non-maîtrise peut gêner la mobilisation de la compétence que l'on cherche à évaluer. La disposition sous forme de tableau peut constituer une compétence écran. De plus pour les exercices B et C, l'intégralité d'un tableau est apte à recevoir les réponses pour toutes les questions posées, ce qui rend compliquée la rectification de ses propres erreurs en cours d'exercice. Il ne s'agit pas de rejeter en bloc ces exercices : de nombreux d'exercices d'évaluation, parce que l'on cherche la plupart du temps à rendre l'élève autonome face à la tâche, comportent des tâches supplémentaires. Toutefois, il nous paraît indispensable de s'engager dans une analyse approfondie des exercices en termes de fonctionnement cognitif afin d'identifier « tout » ce qu'il y a effectivement à faire pour répondre.

Revenons à la présence de tableaux dans des exercices et à leurs effets sur l'activité et les performances des élèves. Bastien et Bastien-Toniazzo (2004, p. 23-26) montrent qu'ils peuvent avoir un effet facilitateur, avec pour résultat essentiel que la difficulté d'un exercice peut résider dans sa modalité de présentation. Dans l'étude rapportée, des élèves de 10-11 ans doivent utiliser l'ordre alphabétique dans un exercice proposé selon deux modalités de présentation et de réponse, l'une dite « tableau », l'autre dite « hexagone », mais avec une liste identique de six mots. Dans la modalité hexagone (la disposition des mots forme un hexagone), il faut tracer, entre les mots, les flèches correspondant à la relation « est avant dans le dictionnaire » ; dans la modalité tableau, il faut faire une croix dans les cases vérifiant cette relation. Contrairement à ce que beaucoup estiment (nous avons interrogé des enseignants sur la difficulté comparée de ces deux modalités), le taux de réussite est bien plus élevé avec la modalité tableau (73 %) qu'il ne l'est avec la modalité hexagone (58 %) ; les omissions, c'est-à-dire les paires de mots pour lesquelles la relation n'a pas été vérifiée, s'élèvent à 11 % et à 39 % respectivement pour les modalités tableau et hexagone. Comme l'expliquent les auteurs (Bastien et Bastien-Toniazzo 2004, p. 26), résoudre cet exercice complexe « demande d'appliquer la relation "est avant" (utilisation d'une connaissance), mais nécessite aussi de considérer et de parcourir l'ensemble des paires de mots (gestion des procédures de réponse). Il se trouve que dans la présentation "tableau", cet ensemble est matérialisé [par les cases], les enfants n'ont alors qu'à le parcourir : ils sont en partie déchargés de la gestion

procédurale de la réponse. En revanche, dans la présentation "hexagone", cet ensemble est virtuel. Les enfants le peuvent ni le construire ni le parcourir exhaustivement ». Ces exercices se distinguent donc du point de vue de leurs caractéristiques contextuelles, lesquelles doivent aussi être analysées du point de vue du fonctionnement cognitif des évalués.

Un troisième enseignant propose (exercice D) :

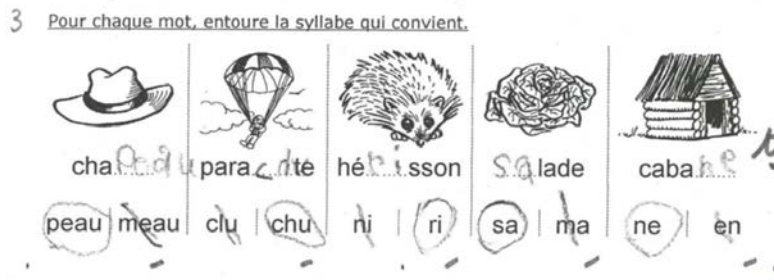


Figure 7. Exercice figurant dans le corpus de la thèse de C. Blanc

Dans les évaluations de CP relatives à la maîtrise de la correspondance entre les lettres et les sons, nous avons identifié un phénomène évaluatif : le recours aux images¹⁶. Alors que seules des lettres et des sons devraient être a priori en jeu, un peu plus d'un exercice sur deux comporte un support imagé ; d'ailleurs, dans une étude consacrée à la VPD, la présence ou l'absence d'un type de support (lettres, sons, images) nous a servi de premier critère pour catégoriser les exercices du corpus visant à évaluer la compétence « connaître les correspondances entre les lettres et les sons ».

Dans l'exercice D, la tâche s'avère relativement complexe pour évaluer une compétence qui serait dite de bas niveau (même si la correspondance entre les lettres et les sons n'est pas toujours facile à maîtriser en début de CP). L'analyse de l'activité nécessaire pour répondre laisse apparaître que l'on s'éloigne de la tâche canonique, valide d'un point de vue psycho-didactique (voir infra). Notons que des élèves ont peut-être déjà rencontré un exercice analogue avant d'être évalués. Mais il arrive toutefois que des enseignants s'emparent, pour le jour de l'évaluation, de ce type d'exercices que l'on trouve abondamment en ligne. Dans le cadre des réflexions sur la validité, en complément de l'approche psycho-didactique, il paraît opportun de s'interroger sur le rapport, ou encore la distance, entre les exercices pour l'évaluation et ceux ayant été travaillés lors de l'apprentissage, autrement dit, il s'agit, en reprenant la terminologie de Scallon (2004), de se préoccuper de la « validité écologique » de l'épreuve.

4.1.4. Le rejet des distracteurs

Du point de vue de leur format, plusieurs des exercices présentés comportent des distracteurs, des réponses à éliminer. L'analyse de l'activité nécessaire à leur rejet fait partie intégrante de l'analyse de la validité d'un point de vue psycho-didactique. Dans ces exercices, l'activité de lecture ne porte pas seulement sur la bonne réponse, mais aussi sur les distracteurs. Les stratégies de réponse et de rejet développées par des élèves de CP peuvent être différentes : par exemple, est-on sûr, qu'une fois la bonne réponse trouvée les élèves ne lisent pas les autres distracteurs ?

Nous avons remarqué que le rejet de certains distracteurs peut potentiellement réclamer une activité bien plus importante que celle permettant de trouver la bonne réponse et faire appel à des compétences qui ne sont pas celles que l'item évalue. Soit un item proposé par Grugeon-Allys et al. (2012) où il s'agit de « cocher la (ou les) expression(s) qui donne(ent) l'aire du rectangle bleu (grand rectangle) » :

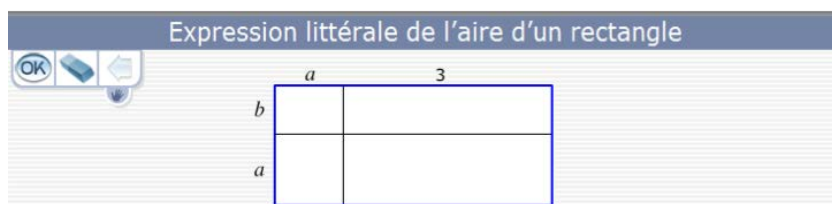


Figure 8. Exercice extrait du test PEPITE - <http://mathenpoche.sesamath.net/>

¹⁶ Nous parlons de « phénomène évaluatif » lorsqu'un « fait évaluatif » est récurrent au sein de la population observée.

Parmi les dix réponses proposées figurent comme bonnes réponses $(a + b)(a + 3)$ et $(a + 3)(a + b)$, ainsi que l'expression développée $ab + 3b + a^2 + 3a$ (en dernière position) qui sont les bonnes réponses, mais aussi des développements erronés comme $3a \times 3b \times a^2 \times ba$. On peut imaginer que le processus de rejet de cette réponse puisse conduire des élèves – qui ne repèrent pas dans la réponse proposée une mauvaise maîtrise de la distributivité dans les calculs algébriques (visible à travers le remplacement « fréquent » des signes « + » par des signes « x ») – à développer $(a + b)(a + 3)$ pour obtenir $a^2 + 3a + ba + 3b$, puis à comparer cette expression avec la réponse proposée.

5. Synthèse et perspectives

5.1. Les deux axes de la validité

Comme le montrent les exemples et analyses précédents, la VPD est une conception qualitative de la validité au sein de laquelle méthodologies et contenus s'articulent. Deux axes d'analyse, tous deux indispensables, pouvant parfois se confondre mais non substituables, permettent de la définir : un premier centré sur les contenus impliqués, un deuxième qui renvoie à ce que les psychologues et ergonomes appellent l'« analyse de la tâche ». Cette distinction a un caractère plutôt formel. Elle a pour but de rappeler la nécessité de ne pas négliger l'un de ces deux axes lors de l'analyse de la VPD d'une tâche. Nous conjugons pour conduire cette analyse des points de vue « méthodologiques » empruntés aux didacticiens, psychologues et ergonomes.

5.1.1. Premier axe

Ce premier axe est centré sur les contenus, sur les connaissances. L'identification et la catégorisation des procédures jouent ici un rôle important. Cet axe renvoie à l'« analyse a priori » des didacticiens que Margolinas (1992, p. 131) présente comme une analyse prévisionnelle, le terme « a priori » signifiant que cette analyse n'est pas dépendante des faits d'expériences, ou des réalisations « effectives » des élèves. Robert (2008, p. 48) présente l'analyse a priori d'un énoncé comme une anticipation portant sur l'activité que des élèves vont pouvoir développer en mobilisant leurs connaissances quand ils sont confrontés à cet énoncé¹⁷. Pour parvenir à répondre au mieux aux trois questions préalables : quel est le savoir en jeu ? quels objectifs sont visés ? quelles connaissances préalables sont nécessaires ? – l'analyse a priori proprement dite porte sur trois points : étude des variables didactiques, études de procédures, étude de l'effet des variables sur les procédures. En dépit du coût de cette analyse, son intérêt est indéniable. Comme le soulignent Julo et Houdebine (1998, p. 79) l'analyse du contenu « doit s'intéresser aussi [aux connaissances] qui sont impliquées de manière implicite dans l'activité et dans la solution du problème posé. Elle doit débusquer celles que l'on ne voit pas au premier coup d'œil ». A titre d'illustration, en s'intéressant à la résolution de problèmes de réinvestissement, Houdement (2011) a montré l'importance de la « qualification » – « qualifier [revenant] à savoir dire (savoir se dire) quelle grandeur "contextualisée" est associée au nombre étudié, que ce nombre soit donné par l'énoncé ou issu d'un calcul » (Houdement 2011, p. 74) –, « connaissance essentielle actuellement invisible dans les curricula » (Houdement 2011, p. 82). Vergnaud (1983, p. 24) ne dissocie pas vraiment deux axes d'analyse comme nous le faisons ici. Il nous enjoint d'entendre par analyse des tâches, auxquelles sont confrontés les élèves dans une situation ou un problème donné, « l'analyse des opérations nécessaires au traitement de la situation et à la solution du problème posé », ces opérations peuvent être diverses et variées (matérielles (faire telle action), perceptives (prendre en compte tel aspect), sociales (interagir ou coopérer avec un tel), etc.) « et elles impliquent des opérations de pensée qui sont nécessairement reliées aux caractéristiques conceptuelles de la situation. L'analyse de ces opérations de pensée, qui représente l'analyse proprement cognitive de la tâche, forme le noyau de l'analyse de la tâche ».

5.1.2. Deuxième axe

Ce second axe correspond largement à la façon dont Julo et Houdebine (1998) présentent l'analyse de la tâche ainsi qu'à l'approche présentée par Bastien et Bastien-Toniazzo (2004). Avec Julo et Houdebine (1998, p. 80), on peut considérer que l'élève se retrouve face à une tâche à réaliser, avec un but à atteindre, et qu'il doit

¹⁷ Robert (2008, p. 48-49) distingue dans le cadre de la réalisation de l'analyse a priori deux principaux niveaux de mise en fonctionnement des connaissances en relation avec des tâches appartenant à deux catégories. D'une part des tâches dites simples et isolées qui amènent à des applications immédiates de connaissances, où seule une connaissance précise est mise en œuvre, sans adaptation. D'autre part, des tâches complexes qui nécessitent des adaptations de connaissances. Robert distingue, pour les tâches complexes, sept grands types d'adaptations, en liens avec les activités des élèves. Pour bâtir sa catégorisation, Robert ne s'intéresse pas uniquement aux tâches et à leurs caractéristiques mais également aux connaissances impliquées et aux activités des élèves en termes d'adaptations des connaissances.

trouver une manière de procéder qui tienne compte des contraintes qui lui sont imposées. Bastien et Bastien-Toniazzo (2004) s'inscrivent, comme Julot d'ailleurs, dans le cadre de la psychologie cognitive développée par Richard. Bastien et Bastien-Toniazzo (2004) s'interrogent sur nombre d'exercices utilisés pour l'évaluation, en procédant à l'analyse des tâches et de l'activité à l'image de ce qui est mis en œuvre en psychologie cognitive et en ergonomie. Pour ces derniers, à juste titre, considérer seulement une analyse conceptuelle ne suffit pas pour interpréter des comportements et des performances d'élèves. Dans leurs travaux, ils montrent qu'il faut être également attentif aux éléments du contexte, aux composantes de la tâche et à ses conditions de réalisation. L'analyse de la VPD, pour être complète, doit comporter un volet que nous nommons « validité liée aux caractéristiques de la tâche », en questionnant notamment la validité liée au contexte matériel. En effet, comme le résume Richard (Bastien et Bastien-Toniazzo, 2004, p. 12-13) :

« des éléments du contexte de la tâche proposée, comme la forme de présentation ou des éléments liés au système de réponse proposé, entrent en interaction avec les connaissances des élèves, ainsi qu'avec les stratégies de réponse [...]. Les choix que font les maîtres dans la présentation de l'exercice engendrent des contraintes qui sont déterminantes pour la réussite ou l'échec. [...] Des changements apparemment minimes du contexte, ne serait-ce que la modification des modalités de réponse choisies, peuvent produire des variations considérables de la fréquence des réponses correctes ».

Il s'agit donc d'analyser des situations concrètes d'évaluation en se centrant sur les points de vue multiples de l'élève confronté à cette situation. Enfin, pour assurer la VPD selon ce second axe, nombre de propositions méthodologiques figurant dans les ouvrages de psychométrie et d'éducatrice et relatives à l'item (conditions de présentation et de réalisation) sont extrêmement pertinentes (voir par exemple Laveault et Grégoire, 2002, p. 24 et suiv.).

5.2. *S'assurer de la validité : une dialectique des fins et des moyens*

S'assurer de la validité d'une épreuve dans une approche psycho-didactique peut s'avérer être une démarche particulièrement coûteuse sur plusieurs plans. Elle repose en effet sur une connaissance approfondie des travaux sur les apprentissages évalués, mais fait également appel à l'expérience d'enseignement et d'évaluation dont disposent les évaluateurs relativement au(x) contenu(s) évalué(s). Mener à bien l'analyse de la VPD nécessite idéalement une analyse de l'activité *effective* des évalués, ce qui impose des observations cliniques de sujets en situation, en amont de l'utilisation des épreuves en situation réelle d'évaluation. En complément, une description statistique des résultats d'une passation impliquant plusieurs tâches et à plus grande échelle peut également contribuer à établir la VPD en considérant conjointement celle-ci et validité « épistémologique ». Il s'agit alors de repérer aussi, à l'échelle d'un test ou de plusieurs épreuves, des incohérences dans les réponses par une analyse des profils de réponses mais aussi d'enrichir les hypothèses diagnostiques comme l'ont montré Grugeon et al. (2012). L'analyse des profils de réponses pourrait à son tour affiner les connaissances sur l'apprentissage du domaine évalué. D'un côté, la multiplicité de ces types et niveaux d'analyses, ainsi que la diversité des prérequis nécessaires pour mener à bien cette analyse de la validité dans sa complétude, pourraient conduire à y renoncer. Si ce type de pratiques est actuellement envisageable dans le cadre de dispositifs d'évaluations externes pour lesquels on dispose du temps suffisant, il pourrait être hors de propos dans le quotidien d'une classe. Mais d'un autre côté, cela encourage à dépasser l'opposition « interne/externe » et à militer pour que se mettent en place des formations initiales et continues à l'évaluation qui adoptent « l'entrée par la didactique » telle que la préconisait Bain dès 1988, afin d'aider les enseignants à répondre, entre autres, aux commandes institutionnelles en matière d'évaluation, en particulier celle qui consiste à proposer une évaluation intégrée aux apprentissages. Ce projet de formation s'apparente à un défi. En effet, si nous pensons que la question de la validité des évaluations est fondamentale, il nous apparaît qu'elle est certainement peu présente au sein des préoccupations professionnelles des enseignants car méconnue, voire ignorée. D'où la nécessité d'intervenir dès la formation initiale. Seule la mise en place de groupes de travail mixtes – composés, d'une part, de chercheurs en didactique, en psychologie cognitive et en évaluation, et, d'autre part, d'enseignants – permettra de développer ce regard critique sur les tâches d'évaluation. L'évaluation peut aussi être un vecteur de transformation des pratiques d'enseignement à partir d'une réflexion sur les tâches, en n'omettant pas de distinguer les tâches pour l'enseignement et pour l'évaluation.

La marque du pluriel pour « APDE » revêt plusieurs significations. D'une part, comme nous l'avons dit, elle signifie que ces approches et la VPD sont pertinentes pour une diversité de « disciplines » et de contenus et que, de ce fait, d'un point de vue pratique, l'analyse de la VPD doit intégrer des éléments divers, en lien avec les spécificités des contenus impliqués dans les évaluations. D'autre part, toujours d'un point de vue pratique, ce pluriel renvoie à une idée importante, plusieurs fois avancée : ces analyses peuvent se faire selon divers degrés de « profondeur », lesquels vont de la « préoccupation » évaluative jusqu'à l'analyse systématique et approfondie de la VPD. Ces préoccupations évaluatives s'incarnent dans certains choix de l'évaluateur qui sont en adéquation avec les principes et finalités présentés dans ce texte, sans que l'évaluateur soit toujours en mesure de les suivre dans leur intégralité. L'analyse systématique et approfondie est quant à elle ambitieuse et reste un

idéal dont il faut connaître l'existence afin d'être en mesure de l'atteindre pour certaines situations d'évaluation. N'oublions pas, par ailleurs, qu'il demeure, pour de nombreuses notions figurant dans les programmes scolaires, des « zones d'ombres », ces notions n'ayant jamais fait l'objet de recherche en didactique (disciplinaire) ou en psychologie (psychologie cognitive du développement ou des apprentissages scolaires). Il serait en effet regrettable de renoncer à se poser des questions sur les contenus et sur ce qui est en jeu dans les tâches au prétexte qu'on ne dispose de l'ensemble des conditions nécessaires à une analyse précise de la validité dans cette perspective. Il nous paraît possible au contraire de s'inspirer des propositions faites ici, entre autres en matière d'analyse a priori et d'analyse de la tâche, sachant, nous insistons, que toute analyse, en particulier du point de vue de sa « profondeur », doit être en adéquation avec les finalités (ou encore avec les objectifs) de l'évaluation. Toutefois, les travaux ayant l'ambition d'une portée plus générale devraient reposer sur une analyse des plus poussées. C'est particulièrement le cas pour les concepteurs d'épreuves externe et les chercheurs. De la diversité de ces situations et contextes, il résulte que s'assurer de la validité dans une approche psycho-didactique n'est pas une procédure standardisée, mais que cela repose immanquablement sur les principes décrits au début de ce texte, et ce dès la conception des épreuves.

Après Cronbach (1980) et De Landsheere (1988), nous soutenons que l'« on n'apportera jamais une preuve parfaite, non équivoque de la validité. Le tout est d'apprécier le degré d'incertitude en la matière » (Cronbach, 1980, cité par De Landsheere, 1988, p. 90) et que « les progrès en matière de validation ne seront probablement pas dus en ordre principal à la sophistication des techniques appliquées, mais bien à une connaissance toujours meilleure des processus » (De Landsheere, 1988, p. 90).

6. Références Bibliographiques¹⁸

Allal, L. (2001). Acquisition et évaluation des compétences en situation scolaire. In J. Dolz et E. Ollagnier (Eds.), *L'énigme de la compétence en éducation*, 77-94. Bruxelles : De Boeck.

Allal, L. (1979). Stratégies d'évaluation formative : conceptions psycho-pédagogiques et modalités d'application. In L. Allal, J. Cardinet et P. Perrenoud (Eds.) *L'évaluation formative dans un enseignement différencié*, 130-157. Berne : Peter Lang.

Allal, L. Bain, D. et Perrenoud, P. (1993). Introduction. In L. Allal, D. Bain et P. Perrenoud (Eds.), *Evaluation formative et didactique du français*, 9-27. Neuchâtel et Paris : Delachaux et Niestlé.

Amigues, R et Zerbato-Poudou, M.-T. (1996). Les pratiques scolaires d'apprentissage et d'évaluation. Paris : Dunod.

Bain, D. (2003). PISA et la lecture : un point de vue de didacticien – Analyse critique de la validité didactique d'une enquête internationale sur la compréhension de l'écrit. *Revue suisse des sciences de l'éducation*, 25/1, 59-78.

Bain, D. (1988). L'évaluation formative fait fausse route : de là, la nécessité de changer de cap. *Mesure et évaluation en éducation*, Vol. 10/4, 23-32.

Bain, D et Schneuwly, B. (1993). Pour une évaluation formative intégrée dans la pédagogie du français : de la nécessité et de l'utilité de modèles de référence. In L. Allal, D. Bain et P. Perrenoud (Eds.), *Evaluation formative et didactique du français*, 51-79. Neuchâtel et Paris : Delachaux et Niestlé.

Bardi, A.-M. et Mégard, A. (2009). L'évaluation scolaire des élèves en France, à un moment charnière de leur histoire. *Mesure et évaluation en éducation*. Vol.32/3, 125-152.

Bastien, C. et Bastien-Toniazzo, M. (2004). *Apprendre à l'école*. Paris : Armand Colin.

Baudelot, C. et Estabiet, R. (2009). *L'élitisme républicain – L'école française à l'épreuve des comparaisons internationales*. Paris : Editions du seuil et de la Républiques des Idées.

Bianco, M. (2010). La compréhension de texte : peut-on l'apprendre et l'enseigner ? In M. Crahay et M. Dutrévis (Eds), *Psychologie des apprentissages scolaires*, 229-256. Bruxelles : De Boeck.

Blanc, C. (2011). Que permettent d'évaluer les évaluations ? Mémoire de master en sciences de l'éducation non publié. Université Paris Descartes.

Bodin, A. (2006). Les mathématiques face aux évaluations nationales et internationales. *Repères IREM n° 65*, 55-89. www.univ-irem.fr/commissions/reperes/consulter/65bodid.pdf

¹⁸ La bibliographie ne comporte pas seulement les références citées dans le texte, mais l'ensemble des documents qui ont permis son écriture. Cette liste est une contribution documentaire sur le thème « évaluations et didactiques ».

Bodin, A. (2005). Ce qui est vraiment évalué par PISA en mathématiques. Ce qui ne l'est pas. Un point de vue français, 27 p. educmath.ens-lyon.fr/Educmath/...math/com_pisa_ff_matheduc.pdf

Bodin, A. (1997). L'évaluation du savoir mathématique – Questions et méthodes. *Recherches en didactique des mathématiques*, Vol. 17/1, 49-96.

Boisnard, D. Houdebine, J. Julo, J. Kerbœuf, M.-P. et Merri, M. (1994). *La proportionnalité et ses problèmes*. Paris : Hachette Education.

Bonniol, J.-J. et Vial, M. (1997). *Les modèles de l'évaluation*. Bruxelles : De Boeck.

Bottani, N et Vrignaud, P. (2005). La France et les évaluations internationales. Rapport établi à la demande du Haut Conseil de l'évaluation de l'école. www.ladocumentationfrancaise.fr > Rapports publics

Braxmeyer, N., Guillaume, J.-C. et Lévy, J.-F. (2004). Les pratiques d'évaluation des enseignants au collège. Note évaluation 12/2004, 04.13 décembre. Vanves : Ministère de l'Éducation nationale - Direction de l'Évaluation et de la Prospective.

Brousseau, G. (2000). Les erreurs des élèves en mathématiques. *Petit x n°57*, 5-30.

Brousseau, G. (1998a). *Théorie des situations didactiques*. Grenoble : La Pensée Sauvage.

Brousseau, G. (1998b). Glossaire de quelques concepts de la théorie des situations didactiques en mathématiques. guy-brousseau.com/wp-content/uploads/2010/09/Glossaire_V5.pdf

Brousseau, G. (1979). L'évaluation et les théories de l'apprentissage en situations scolaires. Accessible sur le site de l'auteur. guy-brousseau.com/wp.../évaluation-et-théorie-de-lapprentissage.doc

Brun, J. (1979). L'évaluation formative dans un enseignement différencié de mathématiques. In L. Allal, J. Cardinet et P. Perrenoud (Eds.) *L'évaluation formative dans un enseignement différencié*, 170-181. Berne : Peter Lang.

Cardinet, J. (1989). Evaluer sans juger. *Revue française de pédagogie n° 88*, 41-52.

Cardinet, J. (1973/1988). L'adaptation des tests aux finalités de l'évaluation. In J. Cardinet *Evaluation scolaire et mesure*, 24-59. Bruxelles : De Boeck.

Carette, V. et Rey, B. (2010). *Savoir enseigner dans le secondaire – Didactique générale*. Bruxelles : De Boeck.

Charnay, R. (1996). *Pourquoi des mathématiques à l'école ?* Paris : ESF.

Chenevotot-Quentin, F., Grugeon-allys, B., Pilet, J. et Delozanne, E. (2012). De la conception à l'usage d'un diagnostic dans une base d'exercices en ligne. In *Actes Espace mathématique francophone GT6*. lutes.upmc.fr/delozanne/Publi/Publi2012/EMF2012-Lingot.pdf

Chevallard, Y. (1986). Vers une analyse didactique des faits d'évaluation. In J.-M. De Ketele (Ed.) *L'évaluation : approche descriptive ou prescriptive ?*, 31-59. Bruxelles : De Boeck.

Crahay, M. (2013). Nécessité et insuffisance d'une psychologie de l'apprentissage pour enseigner les mathématiques. *Education et Formation, e-298-01*, 11-21. ute3.umh.ac.be/revues/index.php?revue=14etpage=3

Crahay, M. (2006). Dangers, incertitudes et incomplétude de la logique de la compétence en éducation. *Revue française de pédagogie n° 154*, 97-110.

De Ketele, J.-M. (2006). La recherche en évaluation : propos synthétiques et prospectifs. *Mesure et évaluation en éducation, Vol. 29/1*, 99-118.

De Ketele, J.-M. et Gérard, F.-M. (2005). La validation des épreuves d'évaluation selon l'approche par les compétences. *Mesure et évaluation en éducation, Vol.28/3*, 1-26.

De Lansheere, G. (1992). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris : Presses universitaires de France.

De Landheere, G. (1976). *Evaluation continue et examens – Précis de docimologie*. Bruxelles : Labor. Paris : Nathan.

De Landsheere, V. (1988). *Faire réussir faire échouer – la compétence minimale et son évaluation*. Paris : Presses universitaires de France.

Delcambre, I. (2007). Contenus d'enseignement et d'apprentissage. In Y. Reuter (Ed.), *Dictionnaire des concepts fondamentaux des didactiques*, 45-51. Bruxelles : De Boeck.

Demonty, I., Fagnant, A. et Dupont, V. (soumis). La mesure des compétences des élèves face à la résolution d'une tâche complexe en algèbre élémentaire : l'apport d'indicateurs issus de recherches centrées sur les difficultés spécifiques des élèves en algèbre.

Dickes, P., Tournois, J., Fieller, A. et Kop, J.-L. (1994). *La psychométrie*. Paris : Presses universitaires de France.

- Ecalte, J. et Magnan, A. (2002). *L'apprentissage de la lecture – Fonctionnement et développement cognitifs*. Paris : Armand Colin.
- Goffard, M. et Weil-Barais, A. (2005). *Enseigner et apprendre les sciences – Recherches et pratiques*. Paris : Armand Colin.
- Grugeon, B. (1997). Conception et exploitation d'une structure d'analyse multidimensionnelle en algèbre élémentaire. *Recherches en didactique des mathématiques*, Vol. 17/2, 167-210.
- Grugeon-Allys, B., Pilet, J., Chenevotot-Quentin, F. et Delozanne, E. (2012). Diagnostic et parcours différenciés d'enseignement en algèbre élémentaire. *Recherches en didactique des mathématiques, Enseignement de l'algèbre, bilan et perspectives, hors série*, 137-162.
- Goasdoué, R., Vantourout, M., Bedoin, D. (à paraître), La correction de dissertations: nouveau regard sur la construction des jugements des évaluateurs. In L. Mottier Lopez & W. Tessaro, *Le Jugement professionnel des enseignants*, Berne : Peter Lang.
- Hadji, C. (1992). *L'évaluation des actions éducatives*. Paris : Presses universitaires de Paris.
- Houdement, C. (2011). Connaissances cachées en résolution de problèmes arithmétiques ordinaires à l'école. *Annales de didactique et de sciences cognitives* Vol. 16, 67-96.
- Johsua, S. et Dupin, J.-J. (1993). *Introduction à la didactique des sciences et des mathématiques*. Paris : Presses universitaires de France.
- Joole, P. (2008). *Comprendre les textes écrits*. Paris : Retz.
- Julo, J. (1995). *Représentation des problèmes et réussite en mathématiques – Un apport de la psychologie cognitive à l'enseignement*. Rennes : Presses universitaires de Rennes.
- Julo, J. et Houdebine, J. (1992). Concevoir de « bonnes » fiches d'activité en mathématiques. *Repères IREM n° 8*, 67-88.
- Laveault, D. et Grégoire, J. (2002). *Introduction aux théories des tests en psychologie et en sciences de l'éducation*. Bruxelles : De Boeck.
- Levain, J-P. (1997). *Faire des maths autrement – Développement cognitif et proportionnalité*. Paris : L'Harmattan.
- Loye, N., Caron, F., Pineault, J., Tessier-Ballairgeon, M., Burney-Vincent, C. et Gagnon, M. (2011). Validité du diagnostic issu d'un mariage entre didactique et mesure sur un test existant. In G. Raïche, K. Paquette-Côté et D. Magis, *Des mécanismes pour assurer la validité pour assurer l'interprétation de la mesure en éducation – Volume 2 – L'évaluation*, 11-29. Canada : Presses universitaires du Québec.
- Margolinas, C. (1992). Eléments pour l'analyse du rôle du maître: les phases de conclusion. *Recherches en didactique des mathématiques* Vol. 12/1, 113-158.
- Marin, B., Crinon, J., Legros, D., et Avel, P. (2007). Lire un texte documentaire scientifique : quels obstacles, quelles aides à la compréhension ? *Revue française de pédagogie n° 160*, 119-131.
- Maurice, J.-J. (1996). Modélisation du savoir-faire de l'enseignant expérimenté : adaptation aux contraintes, anticipations, négociations, pilotage de la classe par les tâches scolaires. Thèse de doctorat, Université Pierre Mendès-France, Grenoble II, Grenoble.
- Maury, S. (2001). Didactique des mathématiques et psychologie cognitive : un regard comparatif sur trois approches psychologiques. *Revue française de pédagogie n° 137*, 85-93.
- Maury, S. (2005). Modélisation du sujet en psychologie et en didactique, dans les travaux intéressant l'enseignement des mathématiques. In M.-H. Salin, P. Clanché, et B. Sarrazy, *Sur la théorie des situations didactiques Questions, réponses, ouvertures. Hommage à Guy Brousseau*, 118-127. Grenoble : La Pensée Sauvage.
- MENJVA/DEGESCO. (2011). Document ressource pour le socle commun dans l'enseignement des mathématiques au collège – Palier 3 (fin de scolarité obligatoire) Compétence 3 – Les principaux éléments de mathématiques et la culture scientifique et technologique. 64 p.
- Mercier, A. (2007). Une question curriculaire de l'enseignement élémentaire des mathématiques : la « résolution de problèmes ». In *Actes du séminaire national : L'enseignement des mathématiques à l'école primaire*, 93-116.
- Morlaix, S. (2009). *Compétences des élèves et dynamique des apprentissages*. Rennes : Presses universitaires de Rennes.
- Mottier Lopez, L. (2007). Régulations interactives situées dans les dynamiques de microculture de classe. *Mesure et évaluation en éducation*, Vol. 30/2, 23-47.
- Ottavi, D. (2009). *De Darwin à Piaget – Pour une histoire de la psychologie de l'enfant*. Paris : CNRS Editions.
- Parrot, F. (2008). Le fonctionnalisme des behaviouristes ; histoire et actualité. In F. Parrot (Ed.), *Les fonctions en psychologie*, 99-116. Wavre : Mardaga.

- Perrenoud, P. (2011). *Quand l'école prétend préparer à la vie ... Développer des compétences ou enseigner d'autres savoirs ?* Bruxelles : De Boeck.
- Perrenoud, P. (1993). Vers des démarches didactiques favorisant une régulation individualisée des apprentissages. In L. Alla I, D. Bain et P. Perrenoud (Eds.), *Evaluation formative et didactique du français*, 31-50. Neuchâtel et Paris : Delachaux et Niestlé.
- Perrenoud, P. (1991). Pour une approche pragmatique de l'évaluation. *Mesure et évaluation en éducation*. Vol. 13/4, 1-33.
- Ravenstein, J. et Sensévy, G. (1994). Le statut de l'erreur dans la relation didactique. *Grand N* n° 53, 83-90.
- Rey, B., Carette, V. Defrance, A et Khan, S. (2003). *Les compétences à l'école – Apprentissage et évaluation*. Bruxelles : De Boeck.
- Riopel, M., Sakr, F., Raïche, G., Potvin, P. et Léocadie Djédjé, V. (2011). Variables de prédiction au niveau de difficulté de tâches d'évaluation comportant des équations du premier degré en mathématiques et en sciences au secondaire. In G. Raïche, K. Paquette-Côté et D. Magis, *Des mécanismes pour assurer la validité pour assurer l'interprétation de la mesure en éducation – Volume 1 – La mesure*, 51-83. Canada : Presses universitaires du Québec.
- Roditi, E. (2003). Régularité et variabilité des pratiques ordinaires d'enseignement. Le cas de la multiplication des nombres décimaux en sixième. *Recherche en Didactique des Mathématiques*, 23/2, 183-216.
- Robert, A. (2008). Une méthodologie pour analyser les activités (possibles) des élèves. In F. Vandebrouck (Ed.) *La classe de mathématiques : activités des élèves et pratiques des enseignants*, 45-56. Toulouse : Octares.
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Bruxelles : De Boeck.
- Scallon, G. (2000). *L'évaluation formative*. Bruxelles : De Boeck.
- Sebaï, N. (2012). Des tâches d'évaluation en mathématiques au livret scolaire : étude qualitative des pratiques de huit enseignants de CM1 et CM2. Thèse de doctorat non-publiée, Université Paris Descartes.
- Vantourout, M. (2004). Etude de l'activité et des compétences de professeurs des écoles et de professeurs de mathématiques dans des situations "simulées" d'évaluation à visée formative en mathématiques, Thèse de doctorat non publiée, Université Paris 5 – René Descartes.
- Vantourout, M. et Goasdoué, R. (2010). Correction de dissertations : analyse de l'activité de professeurs engagés dans une approche par compétences. In *Actes du congrès AREF, Genève, Suisse, septembre 2010*, 12 pages.
- Vantourout, M. et Maury, S. (2006). Quelques résultats relatifs aux connaissances disciplinaires de professeurs stagiaires dans des situations simulées d'évaluation de productions d'élèves en mathématiques. *Revue des sciences de l'éducation* vol. 32, n°3, 759-782. <http://id.erudit.org/iderudit/016285ar>
- Vergnaud, G. (2001). Psychologie du développement cognitif et évaluations des compétence. In G. Figari et M. Achouche (Eds.), *L'activité évaluative réinterrogée*, 43-51. Bruxelles : De Boeck
- Vergnaud, G. (1994). Où en sont les recherches sur l'apprentissage et l'enseignement ? In G. Vergnaud (Ed.), *Apprentissages et didactiques, où en est-on ?*, 5-7.
- Vergnaud, G. (1991). La théorie des champs conceptuels. *Recherche en didactique des mathématiques* Vol.10/2-3, 133-170.
- Vergnaud, G. (1989). Psychologie du développement cognitif et didactique des mathématiques – un exemple : les structures additives. *Petit x* n°22, 51-69.
- Vergnaud, G. (1983). Introduction (didactique et acquisition du concept de volume). *Recherche en didactique des mathématiques* Vol. 4/1, 9-25.
- Vergnaud, G, Brégeon, J.-L., Dossat, L., Huguet, F., Myx, A. et Péault, H. (1997). *Le moniteur de mathématiques*. Paris : Nathan.
- Vrignaud, P. (2006). La mesure de la littéracie dans PISA : la méthodologie mais quelle était la question ? *Revue française de pédagogie* n° 157, 27-41.
- Weil-Barais, A. (1996). *L'homme cognitif*. Paris : Presses universitaires de France.