
## Réseau thématique : Apprentissages Scolaires et Evaluations Externes (ASEE)

# *La gouvernance par les résultats est-elle un mode de régulation de l'école légitime aux yeux des enseignants ? Une enquête qualitative dans 4 systèmes scolaires*

**Gonzague Yerly, Christian Maroy**

A l'image d'autres secteurs publics, la transformation des politiques éducatives suit depuis plusieurs décennies une orientation générale, celle de la «gouvernance par les résultats». Les systèmes éducatifs se dotent de nouveaux mécanismes et outils axés sur l'évaluation de la performance des élèves (contenus standardisés, évaluation externe, contractualisation, reddition de comptes, etc.) qui leur permettent de «piloter» le système et de «réguler» les actions des professionnels sur le terrain, rendus responsables des résultats des élèves (Broadfoot, 2000; Maroy, 2013). Ces politiques de gouvernance par les résultats, appelées aussi politiques d'«accountability» (Mons & Dupriez, 2010) peuvent être réparties en quatre catégories (Maroy & Voisin, 2014) :

Tableau 1 : Typologie des politiques de gouvernance par les résultats (adaptés de Maroy et Voisin, 2014, p. 38)

| | Reddition de comptes « dure » | Reddition de comptes « néo-bureaucratique » | Reddition de comptes et responsabilisation « réflexives » | Responsabilisation « douce » |
|---|---|---|---|---|
| Modes de régulation | Accent sur la reddition de comptes | Accent sur la reddition de comptes | Accent sur la responsabilisation | Accent sur la responsabilisation |
| Outils | Evaluation externe | Evaluation externe et contractualisation | Evaluation externe | Evaluation externe |
| Enjeux sur les acteurs | Forts | Modérés à faibles | Modérés à forts | Modérés à faibles |
| Alignement (outils et actions) | Fort | Fort | Fort | Faibles à modérés |
| Exemples de contextes | Texas, Angleterre | Québec | Ontario, Ecosse | Belgique, France, Suisse |

Si ces réformes sont souvent défendues par les autorités politiques ou scolaires au nom d'une meilleure efficacité ou équité des systèmes, elles font l'objet de critiques des syndicats enseignants ou de chercheurs qui y voient la manifestation d'une perte d'autonomie de l'institution scolaire, d'un déclin de ses visées humanistes ou démocratiques et de l'emprise du monde économique sur l'éducation (Laval & al, 2012). De leur côté, les enseignants sont aujourd'hui confrontés de plus en plus aux outils standardisés de ces politiques, notamment l'évaluation externe. Aussi, par différents mécanismes (responsabilisation, reddition de comptes), il est attendu d'eux qu'ils développent leurs pratiques au regard des «données externes», voire qu'ils contribuent à l'amélioration de la performance du système. Certaines recherches montrent un positionnement complexe des enseignants quant à ces politiques et à leurs outils : notamment une «relative acceptation» de l'évaluation externe (Cattonar, Dumay, & Maroy, 2013), qu'ils perçoivent comme un «mal nécessaire» (Yerly, 2014). De plus, la littérature démontre un usage plutôt faible à modéré des «données externes» par les enseignants. Ceux-ci ne mettraient en place des changements mineurs voire symboliques dans leurs pratiques, ce qui justifie davantage l'argument du

«découplage» (Coburn, 2004) entre politiques éducatives et pratiques en salle de classe. Seul le format des tests leur permet un certain alignement surtout sur les contenus attendus, voire sur certaines méthodes d'évaluation (Rosenwajn & Dumay, 2015; Yerly, soumis). Ce découplage semble particulièrement important dans les contextes où les «enjeux» de ces politiques sont faibles et tablent sur une «responsabilisation douce» (Mons & Dupriez, 2010). Les effets de ces politiques sur les pratiques enseignantes dépendraient surtout des perceptions des enseignants quant aux dispositifs en place (Yerly, 2014).

L'objectif de cette contribution est de comprendre davantage les manières dont les enseignants font sens des politiques de gouvernance par les résultats et des raisons multiples qu'ils mobilisent à cet effet. Plus précisément, nous cherchons à déterminer quels types de légitimité (morale, pragmatique et cognitive - Suchman, 1995) les enseignants accordent ou non aux finalités et aux outils des politiques de gouvernance par les résultats en menant notre analyse dans quatre contextes distincts quant au type de politique de gouvernance par les résultats.

**Méthodologie**

La présente recherche a été menée auprès d'enseignants du primaire (élèves de 6 à 12 ans) de quatre contextes francophones. Il s'agit de contextes où les enjeux des politiques de gouvernance par les résultats restent faibles voire modérés pour les enseignants. Ces contextes représentent trois types de politiques (cf. Tableau 1) :

(1) responsabilisation douce : en Belgique, 1 zone (N=12) et en Suisse, 1 canton (N=16).
(2) reddition de comptes et responsabilisation réflexive : en Ontario, un conseil scolaire (N=14)
(3) reddition de compte néo-bureaucratique : au Québec, une commission scolaire (N=11)

Dans chaque contexte, l'échantillon est formé d'enseignants du primaire en variant les degrés d'enseignement et l'expérience professionnelle. Les enseignants ont été interrogés à l'aide du même guide d'entretien traitant des thèmes suivants : profil professionnel de l'enseignant, profil de l'établissement, les perceptions de l'enseignant quant à la politique de gouvernance et à ses outils, les effets perçus sur ses pratiques. Des informations sur chaque contexte ont été collectées sur chaque contexte par le biais d'entretiens avec des responsables locaux et par une analyse de la documentation officielle. Les 53 entretiens semi-directifs individuels avec les enseignants (de 60 à 80 minutes) ont été enregistrés (audio) et transcrits intégralement pour en faire une analyse de contenu (Miles & Huberman, 2010). Une analyse compréhensive a été menée sur chaque cas, puis sur chaque contexte et enfin les données ont permis une analyse croisée.

**Résultats préliminaires et attendus**

Les premières analyses permettent de constater d'importantes convergences entre les enseignants issus des divers contextes. Notre travail montre plutôt les ambivalences des enseignants à l'égard des principes moraux qui justifient ces politiques (légitimité morale), des théories et présupposés cognitifs qui en sous-tendent la supposée efficacité sur l'amélioration de l'enseignement (légitimité cognitive) et des incidences sur leurs pratiques et contextes de travail concrets (légitimité pragmatique). Ces convergences et cette complexité sont interrogées à partir de l'hypothèse d'une identité professionnelle enseignante partiellement partagée par-delà la diversité des systèmes et des formes prises par ces politiques au plan local.

**Références**

Broadfoot, P. (2000). Un nouveau mode de régulation dans un système décentralisé : l'Etat évaluateur. Revue française de pédagogie, 130(1), 43-55. http://doi.org/10.3406/rfp.2000.1052

Cattonar, B., Dumay, X., & Maroy, C. (2013). Politique d'évaluation externe et recomposition des professionnalités dans l'enseignement primaire : un cas de responsabilisation (accountability) douce. Education et sociétés, n° 32(2), 35-51. http://doi.org/10.3917/es.032.0035

Malet, R. (2011). Un analyseur des formes de redéfinition stratégiques de l'Etat : les politiques

d'imputabilité en direction des enseignants en Angleterre et aux Etats-Unis. In V. Dupriez & N. Mons, Les politiques d'accountability. Du changement institutionnel aux transformations locales (Vol. 5, p. 35-60). Association francophone d'Education comparée. Congrès ADMEE-Europe 2017 4

Maroy, C. (2005). Vers une régulation post-bureaucratique des systèmes d'enseignement en Europe? Les cahiers de Recherche en Education et Formation, (49), 1-30.

Maroy, C., & Voisin, A. (2014). Une typologie des politiques d'accountability en éducation : l'incidence de l'instrumentation et des théories de la régulation. Education comparée -Nouvelle revue, 31-57.

Miles, M., & Huberman, A. M. (2010). Analyse des données qualitatives. Bruxelles : De Boeck Université.

Mons, N., & Dupriez, V. (2010). Les politiques d'accountability. Responsabilisation et formation continue des enseignants. Recherche & formation, n° 65(3), 45-59.

Rosenwajn, E., & Dumay, X. (2015). Les effets de l'évaluation externe sur les pratiques enseignantes : une revue de la littérature. Revue française de pédagogie, n° 189(4), 105-138.

Yerly, G. (soumis). Évaluation en classe et évaluation à grande échelle. Quel est l'impact des épreuves externes sur les pratiques évaluatives des enseignants ? Article soumis à la revue Mesure et évaluation en éducation.

Yerly, G. (2014). Les effets de l'évaluation externe des acquis des élèves sur les pratiques des enseignants. Analyse du regard des enseignants du primaire. (Thèse de doctorat en Sciences de l'éducation). Université de Fribourg, Fribourg. Consulté à l'adresse http://doc.rero.ch/record/256188

*Mots clés : gouvernance par les résultats, légitimité, enseignants du primaire*

## *Motivation et effort des élèves lors des évaluations externes à faibles enjeux : une question de validité et de mesure*

**Christophe Dierendonck, Marianne Milmeister, Christiane Weis, Paul Milmeister**

**Introduction**

Dans le contexte des évaluations externes des acquis scolaires à faibles enjeux pour les élèves, mais à enjeux élevés pour le système (pour le pilotage du système par exemple), il est envisageable qu'un certain nombre d'élèves s'impliquent moins que dans le cas d'un testing à enjeux élevés pour eux. Ce phénomène peut conduire à une sous-estimation plus ou moins forte de la compétence réelle de chaque élève. Si le nombre d'élèves peu impliqués lors des évaluations externes à faibles enjeux est significatif, les résultats globaux observés lors de ces évaluations peuvent s'en trouver également sous-estimés, mettant ainsi en doute la validité des constats dressés. Une question centrale se pose cependant : les élèves apparemment peu impliqués le sont-ils spécifiquement par rapport à la situation de testing à faibles enjeux ou s'agit-il plutôt d'un trait latent que l'on retrouve aussi dans les situations de testing à enjeux élevés, auquel cas les performances observées lors des évaluations externes à faibles enjeux reflètent bel et bien leurs compétences scolaires réelles ?

Des instruments et des méthodologies ont été développés pour tenter d'étudier ce biais éventuel. Selon Rios, Liu et Bridgeman (2014), on peut distinguer trois types d'approches utilisées pour estimer l'implication et l'effort des élèves lors de tests à faibles enjeux pour eux : les mesures auto-rapportées d'effort, les statistiques *person-fit* et l'observation des comportements durant le testing.

Bien que les mesures auto-rapportées de l'effort soient assez faciles à mettre en œuvre tant dans un format papier-crayon que pour des tests sur ordinateur, elles souffrent de plusieurs limites : certains élèves peuvent surévaluer leur effort en raison du phénomène de désirabilité sociale, l'estimation de l'effort peut être influencée par la difficulté perçue du test (Keskpaik & Rocher, 2013) ou par la perception qu'ont les élèves de leurs compétences (Wise & DeMars, 2006) et le niveau d'effort des élèves peut varier au cours du test (Wise & Kong, 2005). Dans les tests dont les résultats sont analysés au travers de la théorie de réponse à l'item, les statistiques *person-fit* (cf. Karabatsos, 2003 ; Meijer & Sijtsma, 2001 ; Wise & Kong, 2005) sont utilisées pour identifier des patrons de réponses aberrants (répondre au hasard par

exemple) et permettent ainsi d'identifier les élèves qui témoignent visiblement d'une faible motivation/implication pour le testing. Lorsque les tests sont réalisés sur ordinateur, il est possible d'enregistrer le comportement des personnes examinées. En particulier, les temps de réponse aux différents items peuvent être enregistrés et donner ainsi la possibilité d'identifier les individus qui, au cours du test, ne prennent pas le temps suffisant pour lire et comprendre en profondeur les questions ou qui n'essaient pas de trouver les réponses.

Notre contribution s'inscrit dans la problématique de la mesure auto-rapportée de l'implication des élèves lors des évaluations externes à faibles enjeux pour eux. La littérature francophone ne propose que peu de travaux à ce sujet (Keskpaik et Rocher, 2012, 2013 ; Dierendonck et al., 2013, 2016). Dans la littérature anglo-saxonne, la mesure de l'implication des élèves est surtout opérationalisée au travers de deux notions : l'effort consenti (*reported effort*) et la perception de l'importance du test (*perceived importance of the test*). Les instruments les plus souvent utilisés dans cette optique sont le *PISA effort thermometer* administré lors des enquêtes PISA 2003, 2006 et 2012, la *Student Opinion Scale* (Sundre et Moore, 2002) et le *Test-taking motivation questionnaire* (Eklöf, 2006). Tous ces instruments sont administrés à l'issue des évaluations.

Notre objectif est de présenter les caractéristiques d'un dispositif de recherche qui pourrait permettre d'examiner la problématique de validité mentionnée plus haut et de répondre à certaines critiques d'instruments existants en rendant compte du développement d'un nouvel instrument de mesure élaboré au départ de deux théories : la théorie *expectancy-value* (Eccles & Wigfield, 2002) et la théorie du comportement planifié (Ajzen, 1991).

## 1. Définition d'un modèle théorique de l'implication des élèves lors d'un test

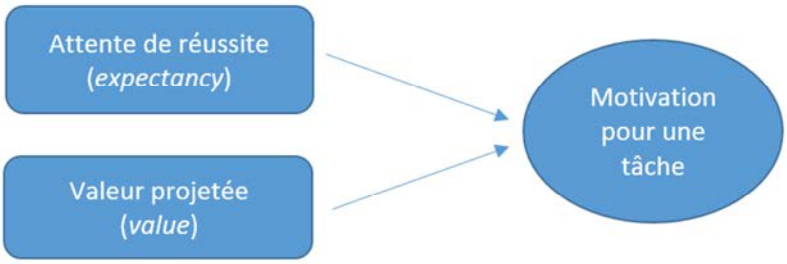
Les deux piliers de notre modèle théorique sont la théorie *expectancy-value* et la théorie du comportement planifié. Au départ, nous avions envisagé de nous baser uniquement sur la théorie *expectancy-value* pour expliquer la motivation des élèves, mais il nous semblait que la théorie du comportement planifié permettait d'apporter des éléments explicatifs manquants, notamment en incluant une dimension sociale. Un autre élément intéressant apporté par la théorie du comportement planifié est que celle-ci tente d'expliquer le comportement par le biais d'une « intention ».

*La théorie expectancy-value*

Il s'agit ici d'une théorie de la motivation développée par John Atkinson (1957) et adaptée au domaine de l'éducation par notamment Eccles et Wigfield (Eccles, Adler, Futterman, Goff, Kaczala, Meece & Midgley, 1983 ; Wigfield, 1994 ; Wigfield & Eccles, 1992 ; Eccles & Wigfield, 2002). L'hypothèse fondamentale de la théorie *expectancy-value* est que la motivation d'un individu explique a) son choix d'une tâche à accomplir, b) sa persistance à cette tâche, ainsi que c) sa performance à cette tâche.

Selon Eccles et Wigfield (2002), la motivation est conditionnée par deux facteurs fondamentaux : les *croyances ou attentes (expectancy)* de l'individu sur sa probabilité de succès à une tâche particulière et la *valorisation (value)* de la tâche en question. La valeur de la tâche (*task value*) est sous-divisée en quatre composantes : l'importance de bien faire la tâche (*attainment value*), la valeur d'intérêt intrinsèque (*instrinsic value*), la valeur d'utilité (*utility value*) et le coût (*cost*). Ce qui distingue l'« *expectancy* » de Eccles et Wigfield d'autres concepts similaires (comme l'auto-efficacité de Bandura par exemple), c'est que celle-ci se réfère à une tâche précise et non à des compétences plus générales.

Figure1 : La théorie *expectancy-value*




Concrètement, l'intérêt de cette théorie pour notre travail réside dans le fait qu'elle se prête fort bien à l'étude du comportement d'élèves face à des tâches différentes, à savoir la passation de tests qui sont soit a) à faibles enjeux ou b) à enjeux élevés. On s'intéressera donc aux attentes de réussite des élèves et à la valeur qu'ils accordent au test pour comprendre leur motivation face à ces deux cas de figure.

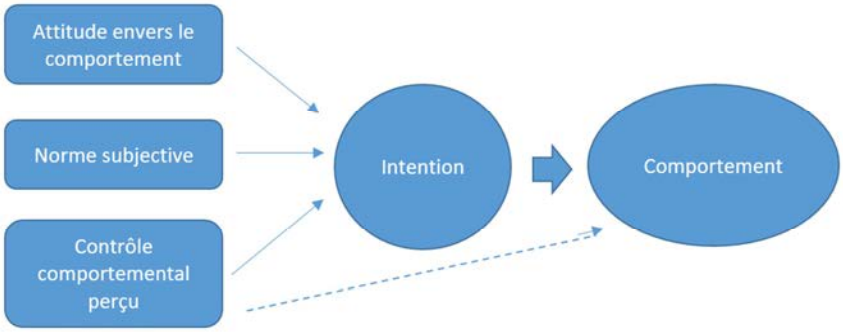
*La théorie du comportement planifié*

La théorie du comportement planifié d'Icek Ajzen est une extension de la théorie de l'action raisonnée, développée par Fishbein et Ajzen (1975). Un élément central dans la théorie du comportement planifié est constitué par l'intention de l'acteur d'adopter ou non un certain comportement. Pour Ajzen (1991), l'intention est le précurseur immédiat de l'action, et elle est déterminée par trois facteurs indépendants : 1) l'attitude envers le comportement en question, 2) la norme subjective, et 3) le contrôle comportemental perçu.

L'attitude envers le comportement fait référence à l'appréciation de l'individu du comportement en question (p.ex. « Je me sens capable de faire ce test sérieusement »). La norme subjective fait référence à la pression sociale perçue par l'individu à adopter le comportement en question ou pas (p.ex. « Mon enseignant attend de moi que je fasse ce test sérieusement »). Enfin, le contrôle comportemental perçu fait référence à la facilité ou difficulté perçue à adopter le comportement en question. Cet indicateur est censé rendre compte aussi bien d'expériences passées que d'obstacles potentiels anticipés (p.ex. « Je me sens capable de rester motivé(e) jusqu'à la fin de ce test »).

L'attitude, la norme subjective, et le contrôle comportemental perçu d'un individu conditionnent son intention d'adopter un comportement ou non. Cette intention est, d'après la théorie du comportement planifié, le meilleur indicateur du comportement effectif adopté, tout en sachant que les limitations posées par le contrôle comportemental perçu pourront avoir une influence directe sur le comportement.

Figure 2 : La théorie du comportement planifié




L'intérêt d'inclure cette théorie dans notre modèle est d'essayer de comprendre le comportement des élèves face aux deux types de test mentionnés plus haut sous un deuxième angle théorique. La théorie du comportement planifié est axée sur la prédiction d'un comportement via l'intention. Concrètement, nous avons défini le comportement à expliquer comme l'investissement effectif lors du test et son précurseur immédiat comme l'intention de s'investir. Nous mesurerons donc l'attitude, les normes subjectives et le

contrôle comportemental perçu des élèves concernant leur investissement en situation de testing à enjeux faibles et à enjeux élevés pour eux.

*Combinaison des deux modèles théoriques et variables additionnelles*

Nous avons tenté de combiner les deux théories citées en une seule modélisation (figure 3). Effectivement, comme exposé plus haut, non seulement la théorie du comportement planifié apporte un élément social qui manque à la théorie *expectancy-value*, mais encore il s'agit de deux approches théoriques différentes, où chacune met l'emphase sur d'autres éléments explicatifs centraux :

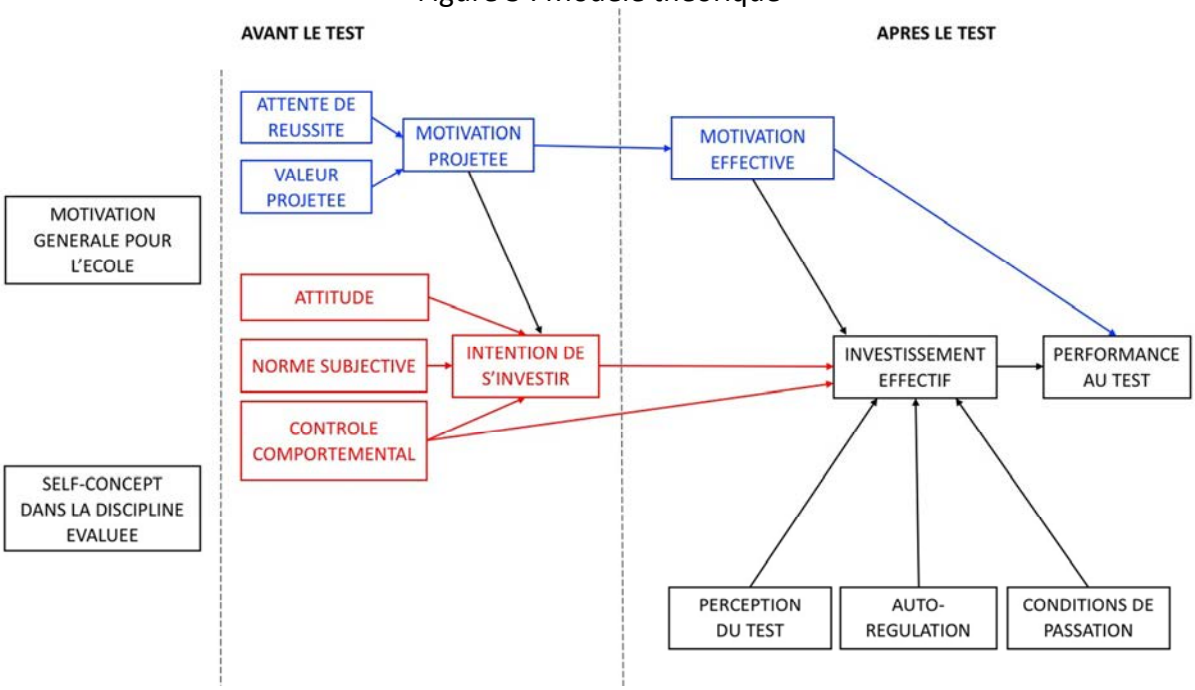
- alors que la théorie *expectancy-value* essaie d'expliquer la performance à une tâche, la théorie du comportement planifié veut expliquer un comportement ;
- alors que la théorie *expectancy-value* se focalise sur la motivation comme élément explicatif primordial, la théorie du comportement planifié est centrée sur l'intention comportementale de l'acteur.

Il est effectivement pertinent de se demander dans quelle mesure ces deux indicateurs se distinguent (ou se rejoignent) et quelles sont leurs capacités prédictives relatives. Dans notre modèle, ce sont donc aussi bien la motivation de l'élève à la tâche que son intention de s'investir qui influencent ce que nous appelons l'investissement effectif. Ce dernier est alors prédicteur de la performance au test.

Une originalité de notre instrument de mesure est le fait qu'il comprend deux temps de passation ; il y a une prise de données avant le test et une après le test. Ceci permet au niveau théorique de faire par exemple la distinction entre la motivation projetée pour le test et la motivation effective lors du test. De la même façon, nous distinguons l'intention de s'investir de l'investissement effectif lors du test.

Nous souhaitions aussi distinguer, au sein des données motivationnelles, le trait latent (la motivation générale pour l'école) du trait spécifique à la situation (la motivation pour le test). Sans cette information, il est en effet impossible de savoir si les élèves qui se disent peu motivés lors d'un test à faibles enjeux pour eux le sont uniquement pour le test en question ou s'ils le sont aussi pour l'école en général. Nous avons également fait l'hypothèse que le *self-concept* des élèves dans la matière scolaire concernée par le test (ici, les mathématiques) pouvait avoir une influence sur les variables de notre modèle. Concrètement, nous avons appréhendé la motivation générale des élèves pour l'école et leur self-concept en mathématiques au travers d'items repris et adaptés de la version francophone du *Self-description Questionnaire-I* de Marsh (1988) validée par un des auteurs de cette contribution (Dierendonck, 2008).

Figure 3 : Modèle théorique



Dans ce modèle explicatif de la performance observée au test (à enjeux élevés ou faibles), nous postulons aussi qu'à côté de la motivation et de l'intention de s'investir, l'investissement effectif des élèves peut être conditionné par trois éléments supplémentaires : la perception effective du test (Le test est-il difficile, intéressant, utile,… ?), les comportements d'auto-régulation que l'élève adopte durant le test (p.ex. relire ses réponses) et les conditions de passation (Y avait-il du bruit dans la classe ? Les autres élèves étaient-ils sérieux ?).

**2. Elaboration de l'instrument de mesure**

Pour élaborer le nouvel instrument de mesure de l'implication des élèves lors d'un test, nous avons adapté la démarche de Ndinga et Frenette (2010), inspirée de De Vellis (2003), pour tenir compte de notre contexte scolaire multilingue. Nous avons observé cinq étapes : (1) détermination de l'objet de mesure à la lumière de la théorie sous-jacente, (2) génération des items en français, (3) détermination du format de mesure, (4) vérification de la clarté des questions et traduction des items en allemand et (5) prétest de l'instrument. Le processus de validation de l'instrument (analyse d'items et vérification de la structure factorielle de l'instrument) auprès d'un large échantillon doit encore être conduit.

*Etapes 1 à 3 : Construction du questionnaire*

Pour couvrir les dimensions de notre modèle théorique, 57 items ont été formulés. Le tableau 1 détaille le nombre d'items par dimension. Six items (repris entre parenthèses dans le tableau) sont spécifiques à la situation de testing à faibles enjeux.

Tableau 1 : Dimensions et items de l'instrument – Version initiale

| *Avant l'évaluation* | |
|---|---|
| Dimensions | Nb items |
| Motivation générale pour l'école | 3 |
| Self-concept en math | 3 |
| Attente de réussite | 3 |
| Valeur projetée | 5 |
| Motivation projetée | 3 |
| Attitude | 3 |
| Norme subjective | 3 |
| Contrôle comportemental | 4 |
| Intention de s'investir | 3 |

| *Après l'évaluation* | |
|---|---|
| Dimensions | Nb items |
| Motivation effective | 3 (+3) |
| Perception du test | 8 |
| Auto-régulation | 3 |
| Conditions de passation | 4 |
| Investissement effectif | 3 (+3) |

Les élèves devaient se positionner par rapport à chaque item sur une échelle de type Likert à six modalités de réponse (trois modalités à valence négative et trois modalités à valence positive), pouvant varier d'un item à l'autre (ex : 0 = pas du tout d'accord à 5 = tout à fait d'accord ou 0 = pas du tout motivé à 5 = tout à fait motivé). En complément des données récoltées au niveau des élèves, nous avons demandé aux enseignants de nous fournir, pour chaque élève, une appréciation notée de 0 à 10 rendant compte, d'une part, du degré de motivation générale pour l'école et, d'autre part, du niveau général en mathématiques.

*Etape 4 : Traduction des items et vérification de la clarté des items*
La version française de l'instrument a été relue, discutée et validée par les quatre auteurs. Les items ont ensuite été traduits en allemand par un auteur. Les traductions ont été vérifiées par les autres auteurs. Sur la base d'un échange entre les auteurs, plusieurs clarifications et améliorations ont été apportées aux traductions. La version allemande de l'instrument n'a pas, à ce stade, fait l'objet d'une nouvelle traduction en français pour s'assurer de l'équivalence des deux versions de l'instrument. Ce processus est cependant prévu avec la version finalisée du questionnaire (en annexe) avant l'étude à large échelle.

*Etape 5 : Prétest de l'instrument*
Le prétest de l'instrument a eu lieu en décembre 2016 auprès de 99 élèves de VI$^{e}$ (ce qui correspond à la deuxième année de l'enseignement secondaire ou au grade 8) répartis dans quatre classes d'un établissement d'enseignement secondaire classique. L'administration du questionnaire a été réalisée par le premier auteur (2 classes) et le quatrième auteur (2 classes).
Le prétest comprenait deux temps : (temps 1) la version « enjeux élevés » de l'instrument de mesure de l'implication des élèves a été administrée avant (30 items) et après (21 items) une évaluation en mathématiques préparée par l'enseignant et comptant pour la note trimestrielle et (temps 2) la version « enjeux faibles » de l'instrument de mesure a été administrée avant (30 items) et après (27 items) un test de compétence en mathématiques préparé par nos soins et présenté aux élèves comme ne comptant pas pour des points.

**3. Résultats du prétest**

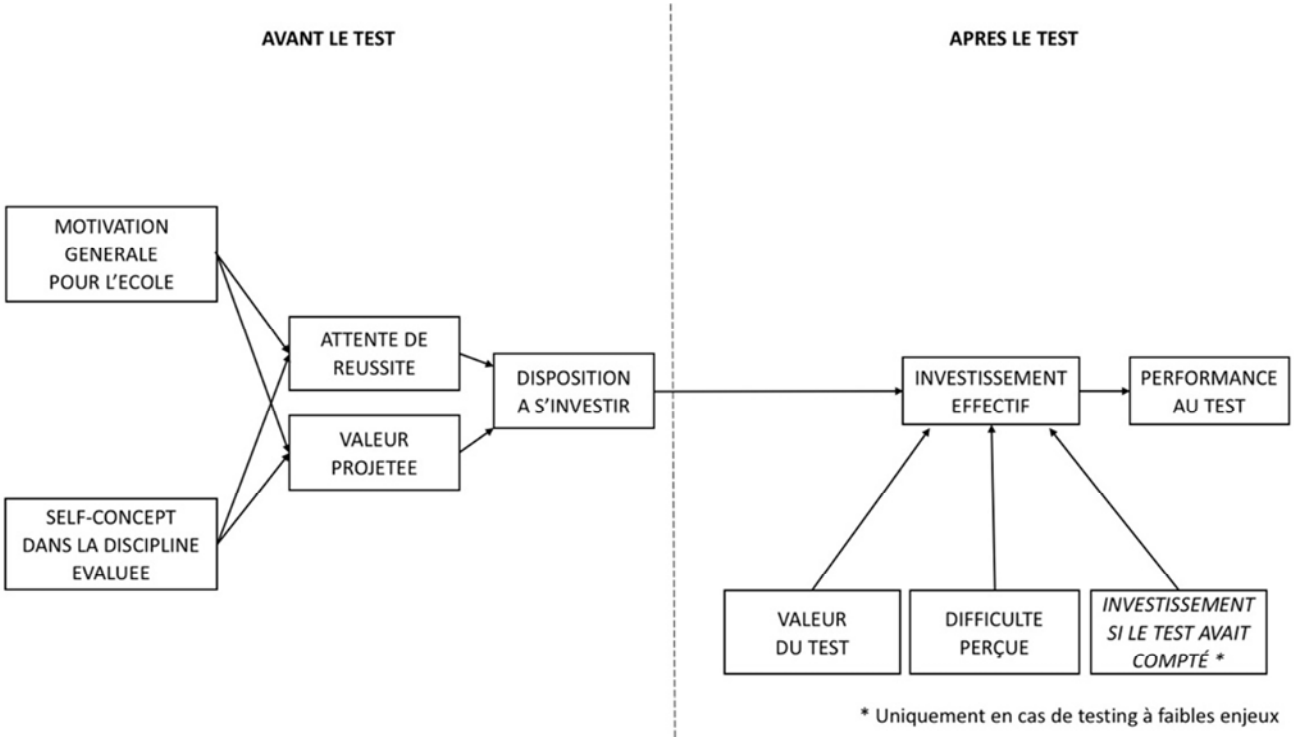
*Repérage des dimensions problématiques*
Nous avons considéré qu'une dimension était problématique à partir d'un coefficient de consistance interne ou alpha de Cronbach sensiblement inférieur à 0,70. Ce fut le cas pour les dimensions suivantes : contrôle comportemental (1$^{e}$ partie du questionnaire), perception du test (2$^{e}$ partie du questionnaire), auto-régulation (2$^{e}$ partie du questionnaire) et conditions de passation (2$^{e}$ partie du questionnaire).

*Repérage des items problématiques*
En parallèle du repérage des dimensions problématiques, nous avons examiné, pour chaque dimension théorique, l'existence éventuelle d'items problématiques sur la base du respect de deux principes édictés par Ndinga et Frenette (2010), à savoir : (1) la corrélation entre un item et les autres items censés appartenir à la même dimension ne peut être trop faible ($r < 0,20$) et (2) la corrélation entre un item et les items d'une autre dimension ne peut être trop élevée ($r > 0,60$).
Cette analyse a permis de mettre en lumière des corrélations trop élevées entre tous les items des dimensions relatives à la motivation projetée pour le test, d'une part, et à l'intention de s'investir d'autre part. Ce fut le cas aussi entre tous les items des dimensions « attente de réussite » et « contrôle comportemental », ainsi qu'entre les items de la dimension « valeur projetée » d'une part et les items des dimensions « attitude » et « norme subjective » d'autre part.
En outre, certains items de la dimension « perception du test » étaient fortement liés entre eux. Ils ont pu être rassemblés en deux dimensions signifiantes et consistantes : (1) difficulté perçue du test (3 items) et valeur perçue du test (4 items).

*Réajustement du modèle et de l'instrument*
Les constats précédents nous ont conduits à renoncer à la distinction théorique opérée entre la motivation (projetée et effective) pour le test (théorie *expectancy-value*) et le comportement (intention de s'investir et investissement effectif) lors du test (théorie du comportement planifié). Nous avons donc abandonné l'idée de tester un modèle combinant parfaitement les deux théories, puisque la distinction entre motivation et intention de s'investir s'est avérée inopérante sur le plan empirique. La théorie *expectancy-value* a été retenue comme théorie de référence principale, tout en gardant la spécificité de la théorie du

comportement planifié, à savoir l'existence d'une variable intermédiaire entre les aspects motivationnels et le comportement visé. Concrètement, nous avons considéré une seule dimension pour mesurer, avant le test, la motivation des élèves à s'investir (disposition à s'investir) et une seule dimension pour appréhender leur implication réelle (investissement effectif), mesurée après le test.
Le modèle théorique a donc été revu (figure 4) et l'instrument a subi une réduction en termes de nombre de dimensions et de nombre d'items (voir annexe 1).

Figure 4 : Modèle ajusté




Les tableaux 2 et 3 présentent les caractéristiques de l'instrument retenu en renseignant l'indice de consistance interne de chaque dimension, respectivement pour la situation de testing à enjeux élevés et pour la situation de testing à enjeux faibles pour les élèves. A la lecture de ces tableaux 2 et 3, on constate qu'à l'exception des dimensions « disposition à s'investir » dans la situation à enjeux élevés (alpha de .66) et « investissement si le test avait compté » dans la situation à enjeux faibles (alpha de .63), toutes les dimensions présentent des indices satisfaisants (>.70) voire très satisfaisants (>.80), ce qui autorise la construction d'indicateurs synthétiques.

*Statistiques descriptives*
La partie gauche du graphique 1 présente les distributions, sous la forme de boîtes à moustache, des indicateurs calculés pour chaque dimension, respectivement pour la situation d'évaluation à enjeux faibles et la situation d'évaluation à enjeux élevés. La partie droite du graphique renseigne l'intervalle de confiance calculé autour de chaque indicateur moyen afin de mettre en évidence les différences statistiquement significatives au seuil de 5%. Ces différences significatives sont marquées d'un astérisque.

Tableau 2 : Dimensions et items de l'instrument – Version modifiée – Situation à enjeux élevés

| *Avant l'évaluation* | | | |
|---|---|---|---|
| Dimensions | Nb items | Alpha | N |
| Motivation générale pour l'école | 3 | .82 | 98 |
| Self-concept en mathématiques | 3 | .92 | 97 |
| Attente de réussite | 3 | .84 | 97 |
| Valeur projetée | 4 | .71 | 96 |
| Disposition à s'investir | 6 | .66 | 98 |

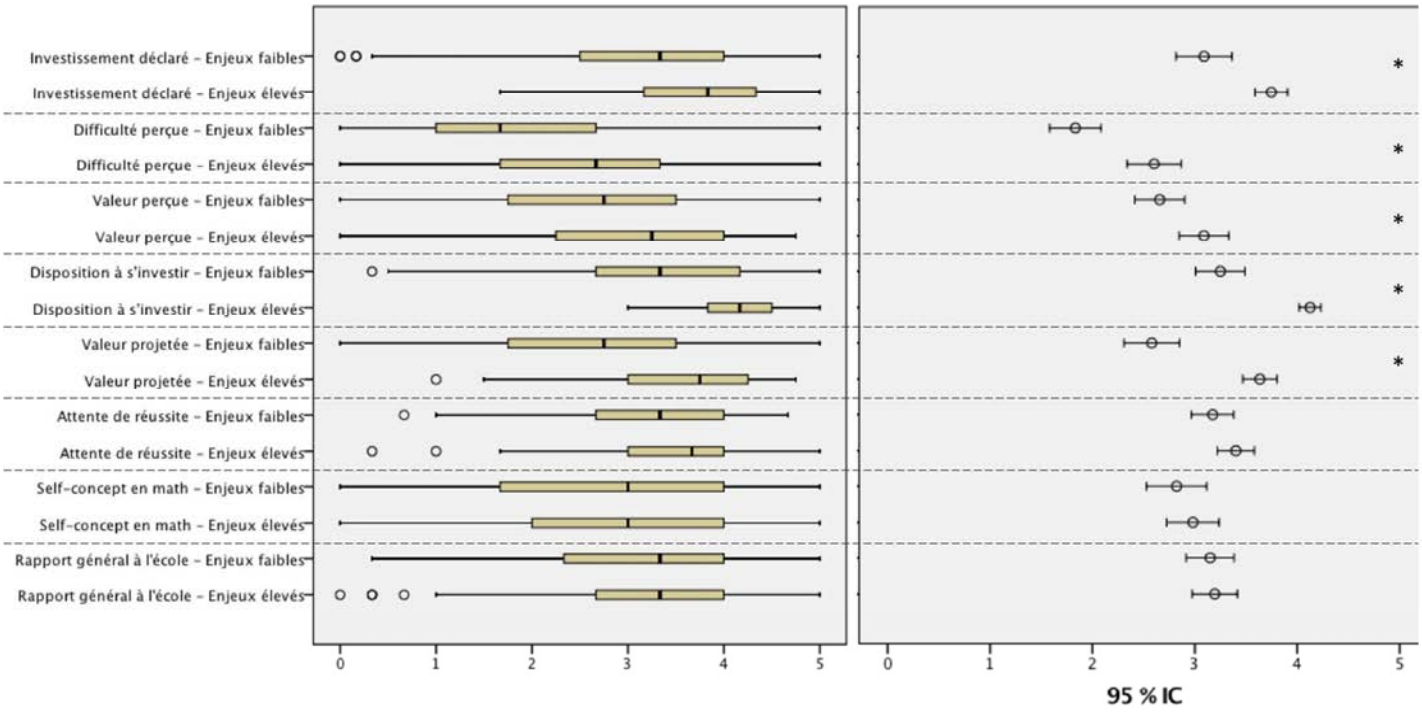
| *Après l'évaluation* | | | |
|---|---|---|---|
| Dimensions | Nb items | Alpha | N |
| Valeur du test | 4 | .84 | 89 |
| Difficulté perçue | 3 | .88 | 92 |
| Investissement effectif | 6 | .82 | 96 |

Tableau 3 : Dimensions et items de l'instrument – Version modifiée – Situation à enjeux faibles

| *Avant l'évaluation* | | | |
|---|---|---|---|
| Dimensions | Nb items | Alpha | N |
| Motivation générale pour l'école | 3 | .80 | 91 |
| Self-concept en mathématiques | 3 | .91 | 89 |
| Attente de réussite | 3 | .81 | 89 |
| Valeur projetée | 4 | .84 | 89 |
| Disposition à s'investir | 6 | .94 | 91 |

| *Après l'évaluation* | | | |
|---|---|---|---|
| Dimensions | Nb items | Alpha | N |
| Valeur du test | 4 | .76 | 92 |
| Difficulté perçue | 3 | .77 | 92 |
| Investissement effectif | 6 | .93 | 85 |
| Investissement si le test avait compté | 6 | .63 | 84 |

Graphique 1 : Statistiques descriptives (dispersion et moyenne)
pour chaque variable du modèle en fonction de la situation de testing (enjeux faibles vs enjeux élevés)




*Liens entre les indicateurs et la performance observée aux évaluations*

Compte tenu des faibles effectifs de l'échantillon, il serait hasardeux de se lancer dans des modélisations statistiques élaborées. Nous nous sommes donc limités à calculer les corrélations au niveau de l'échantillon. Nous présentons ici quelques constats que nous estimons les plus intéressants. Le tableau 4 présente la matrice de corrélation entre les différentes dimensions. Il faut cependant insister sur le fait que la variable de performance dans la situation à enjeux élevés a été obtenue à partir de quatre évaluations incomparables puisque différentes sur plusieurs aspects (contenu évalué, difficulté du devoir, importance pour le trimestre, degré de sévérité lors de la notation, …). Les corrélations relatives à cette variable sont donc à considérer avec la plus grande précaution (elles sont reprises en grisé dans le tableau).

Parmi les relations intéressantes, on évoquera tout d'abord l'indice de corrélation statistiquement significatif, mais de faible ampleur (.29), entre le score observé au test sans enjeux et le niveau général en mathématiques rapporté par les enseignants. Le graphique 2 permet de comprendre la faiblesse du lien dégagé puisqu'on observe que les élèves dont le niveau en mathématiques est jugé moyen (4 à 6) par les enseignants obtiennent des scores au test très variables (entre 0 et 8 sur 10). On remarque aussi l'existence d'élèves jugés bons ou très bons (7 ou plus) par les enseignants qui obtiennent un score au test assez faible (5 ou moins). A l'inverse, quelques élèves jugés faibles ou très faibles en mathématiques obtiennent des scores satisfaisants (5 ou plus) au test sans enjeux. Pour expliquer ces constats, plusieurs hypothèses sont envisageables. Ainsi, les élèves jugés bons par leur enseignant mais qui n'ont pas forcément bien performé lors du test sans enjeux pour eux ont sans doute fait preuve d'un manque d'implication, mais il est également possible qu'ils aient été déstabilisés par cette situation de testing face à laquelle aucune préparation préalable n'était possible. A l'opposé, les quelques élèves qui ont obtenu un score au test sans enjeux largement supérieur aux jugements de l'enseignant (c'est surtout vrai pour l'élève de la classe A qui obtient un score de 7/10 au test alors que son niveau général en mathématiques est considéré comme très faible par l'enseignant) ont soit démontré une meilleure aptitude à faire face à une situation de testing où aucune préparation n'était nécessaire, soit voulu prouver de quoi ils étaient vraiment capables lorsque leurs compétences sont évaluées de manière externe à la classe et à

l'enseignant.

Graphique 2 : Score observé au test sans enjeux
en fonction du niveau général en mathématiques estimé par les enseignants

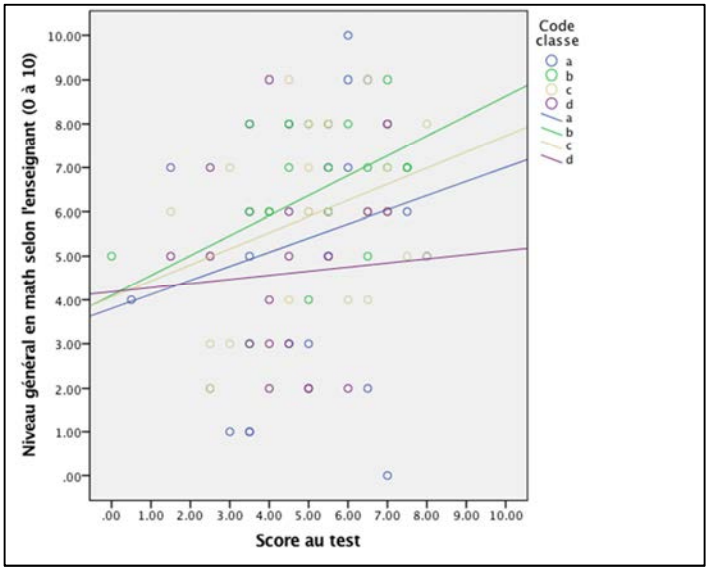



Tableau 4 : Matrice des corrélations de Pearson entre les différentes dimensions

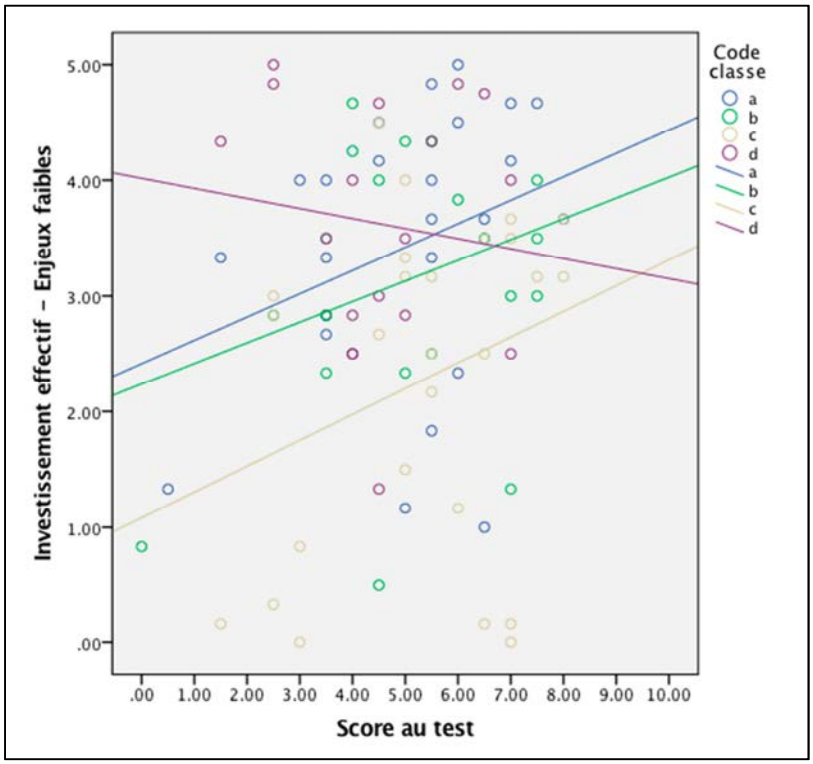
| Dimensi | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Score | 1.00 | | | | | | | | | | | | | | | | | | | |
| 2. Score | .24* | 1.00 | | | | | | | | | | | | | | | | | | |
| 3. Motivati | .35** | .11 | 1.00 | | | | | | | | | | | | | | | | | |
| 4. Niveau | .66** | .29** | .69** | 1.0 0 | | | | | | | | | | | | | | | | |
| ***Situatio*** | | | | | | | | | | | | | | | | | | | | |
| 5. Motivati | .11 | .03 | .34** | .27** | 1.0 0 | | | | | | | | | | | | | | | |
| 6. Self- | .35** | .22* | .3 | .42 | .54 | 1.0 | | | | | | | | | | | | | | |
| 7. Valeur | .11 | .03 | - | .08 | .36 | .45 | 1.0 | | | | | | | | | | | | | |
| 8. | .24* | .11 | .2 | .32 | .34 | .55 | .41 | 1.0 | | | | | | | | | | | | |
| 9. | .31** | .13 | .1 | .22 | .33 | .37 | .53 | .56 | 1.0 | | | | | | | | | | | |
| 10. | .25* | -.06 | - | .06 | .43 | .39 | .69 | .42 | .47 | 1.0 | | | | | | | | | | |
| 11. | -.31** | .00 | - | - | - | - | - | - | - | - | 1.0 | | | | | | | | | |
| 12. | .29** | .02 | .1 | .14 | .39 | .35 | .40 | .48 | .71 | .57 | - | 1.0 | | | | | | | | |
| ***Situatio*** | | | | | | | | | | | | | | | | | | | | |
| 13. Motivati | .14 | .01 | .33** | .26* | .90** | .44** | .29** | .23* | .34** | .36** | -.15 | .47** | 1.0 0 | | | | | | | |
| 14. Self- | .37** | .18 | .2 | .35 | .47 | .88 | .37 | .47 | .41 | .39 | - | .42 | .42 | 1.0 | | | | | | |
| 15. | .08 | -.03 | - | .01 | .54 | .35 | .49 | .18 | .37 | .56 | - | .42 | .55 | .45 | 1.0 | | | | | |
| 16. | .34** | .01 | .2 | .28 | .44 | .50 | .26 | .60 | .52 | .36 | - | .52 | .44 | .60 | .50 | 1.0 | | | | |
| 17. | .17 | .00 | .0 | .12 | .59 | .46 | .51 | .29 | .55 | .52 | - | .55 | .63 | .55 | .83 | .64 | 1.0 | | | |
| 18. | .10 | -.03 | - | - | .40 | .21 | .38 | .19 | .34 | .46 | - | .35 | .43 | .34 | .85 | .42 | .75 | 1.0 | | |
| 19. | -.15 | -.35** | - | - | .07 | - | .01 | - | - | - | .12 | - | .08 | - | .08 | - | .04 | .14 | 1.0 | |
| 20. | .28** | .16 | .0 | .12 | .54 | .38 | .41 | .30 | .46 | .50 | - | .56 | .56 | .49 | .76 | .61 | .85 | .78 | .00 | 1.0 |
| 21. Investiss | -.12 | -.01 | -.0 | .01 | -.02 | -.04 | .01 | -.01 | .11 | .01 | -.05 | -.01 | -.01 | -.04 | -.17 | .05 | -.15 | -.21 | .03 | -.24 |

* La corrélation est significative au niveau 0,05 (bilatéral).
** La corrélation est significative au niveau 0,01 (bilatéral).

On peut s'intéresser aussi au lien qui existe entre le score observé au test sans enjeux et la difficulté de ce test perçue par les élèves. Avec un indice de corrélation négatif (-.35), cela révèle la tendance suivante :

moins les élèves jugent le test difficile, plus leur score au test sans enjeux est élevé. Cette relation s'explique en partie logiquement car on peut supposer que plus un élève est compétent, moins il trouvera un test difficile et moins il reconnaitra qu'il a fourni des efforts et inversement.
Au niveau de l'échantillon, la corrélation entre le score observé au test sans enjeux et le degré d'investissement effectif par les élèves n'est pas statistiquement significative, ce qui irait à l'encontre de notre hypothèse selon laquelle la performance à un test dépend non seulement de la compétence de l'élève mais également de son niveau d'implication ou du sérieux avec lequel il fait le test. Le graphique 3 permet de comprendre pourquoi la corrélation calculée au niveau de l'échantillon est si faible. En effet, on s'aperçoit tout d'abord que la relation qui unit le score et l'investissement est négative dans la classe d (en moyenne, les scores les plus élevés au test sont associés à des niveaux d'investissement plus faibles), alors qu'elle est positive dans les classes a, b et c. On relève aussi l'existence d'un certain nombre d'élèves qui déclarent s'être très peu investis, alors que leur score au test s'avère satisfaisant. Enfin, on rappellera l'influence des comportements de réponse extrêmes qui peuvent biaiser les indices de corrélation.

Graphique 3 : Score observé au test sans enjeux
en fonction du degré d'investissement effectif par les élèves




Sur la base du constat précédent mettant en évidence que le sens des relations entre dimensions peut varier d'une classe à l'autre, nous avons calculé les indices de corrélation en considérant séparément les données de chaque classe. Compte tenu du nombre d'élèves par classe, ces indices sont le plus souvent statistiquement non significatifs, mais ils donnent au moins une idée sur le sens des relations. Le tableau 5 présente ces indices pour les dimensions relatives à l'évaluation à enjeux élevés. Le tableau 6 concerne quant à lui la situation de testing sans enjeux.

Tableau 5 : Indices de corrélation entre le score obtenu à l'évaluation à enjeux élevés et les dimensions étudiées en fonction de la classe

| | Classe a | Classe b | Classe c | Classe d |
|---|---|---|---|---|
| Motivation générale pour l'école | 0.29 | -0.06 | -0.04 | 0.37 |
| Self-concept en mathématiques | 0.40* | 0.30 | 0.38 | 0.24 |
| Valeur projetée | -0.13 | 0.35 | 0.02 | 0.00 |
| Attente de réussite | 0.20 | 0.55** | -0.05 | 0.29 |
| Disposition à s'investir | 0.24 | 0.54** | 0.27 | 0.03 |
| Valeur perçue | 0.15 | 0.37 | 0.19 | 0.18 |
| Difficulté perçue | -0.55* | -0.48* | -0.22 | -0.40 |
| Investissement effectif | 0.41* | 0.39 | 0.09 | 0.06 |
| Motivation générale pour l'école selon l'enseignant | 0.83** | 0.70** | 0.21 | 0.26 |
| Niveau général en math | 0.88** | 0.91** | 0.57** | 0.69** |

* La corrélation est significative au niveau 0,05 (bilatéral).
** La corrélation est significative au niveau 0,01 (bilatéral).

Tableau 6 : Indices de corrélation entre le score au test sans enjeux et les dimensions étudiées en fonction de la classe

| | a | b | c | d |
|---|---|---|---|---|
| Motivation générale pour | 0.07 | -0.17 | 0.16 | -0.19 |
| Self-concept en math | 0.17 | 0.50* | 0.31 | -0.23 |
| Valeur projetée | 0.11 | 0.05 | 0.01 | -0.12 |
| Attente de réussite | 0.24 | 0.15 | -0.04 | -0.31 |
| Disposition à s'investir | 0.18 | 0.05 | 0.11 | -0.23 |
| Valeur perçue | -0.01 | 0.06 | 0.14 | -0.18 |
| Difficulté perçue | -0.32 | -0.41 | -0.35 | -0.33 |
| Investissement effectif | 0.32 | 0.29 | 0.29 | -0.12 |
| Motivation générale pour | 0.14 | 0.38 | 0.48 | -0.24 |
| Niveau général en math | 0.20 | 0.49* | 0.36 | 0.07 |

* La corrélation est significative au niveau 0,05 (bilatéral).
** La corrélation est significative au niveau 0,01 (bilatéral).

**Conclusion et perspectives**

L'objectif de cette communication était de rendre compte du prétest d'un modèle théorique et d'un instrument de mesure de l'implication des élèves lors d'une évaluation de leurs acquis scolaires, qu'elle soit à enjeux élevés ou à enjeux faibles pour leur parcours scolaire.

D'un modèle théorique tentant de combiner les apports de la théorie *expectancy-value* et de la théorie du comportement planifié, nous avons abouti à un modèle simplifié incluant, selon la situation de testing considérée, huit ou neuf dimensions explicatives de la performance lors d'une évaluation. A la lumière des données recueillies dans quatre classes, ces dimensions semblent être appréhendées au travers d'échelles de mesure témoignant de qualités psychométriques satisfaisantes.

A l'inverse des mesures classiques de l'implication des élèves basées exclusivement sur la notion d'effort et la motivation pour le test, nous proposons une échelle d'investissement qui inclut différents aspects : motivation à réussir, concentration, travail sérieux et effort (cf. annexe).

Pour pouvoir examiner quelle est l'influence réelle de l'investissement déclaré par les élèves sur la performance observée aux évaluations à enjeux faibles pour eux, il convient à présent de mettre à l'épreuve le modèle théorique et l'instrument de mesure auprès d'un échantillon plus large d'élèves. Mais auparavant, il s'agit d'améliorer le dispositif de recherche car le prétest a permis de pointer plusieurs biais potentiels liés notamment à la fiabilité et à la comparabilité des indicateurs de compétence en situation d'évaluation à enjeux élevés (puisque les évaluations différaient selon la classe et que nous ne pouvions pas contrôler le degré de sévérité des enseignants) et au caractère artificiel de la situation de testing à faibles enjeux. Sur ce dernier point, l'idéal serait de pouvoir organiser une collecte de données au sein des dispositifs officiels d'évaluation externe des acquis scolaires, qu'ils soient à enjeux élevés ou à enjeux faibles pour les élèves.

**Références**


Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211.

Atkinson, J.W. (1957). Motivational determinants of risk taking behavior. *Psychol. Rev. 64*: 359-372.

De Vellis, R. F. (2003). Scale development: Theory and Applications (2e éd.). Thousand Oaks: Sage Publications.

Dierendonck, C. (2008). Validation psychométrique d'un questionnaire francophone de description de soi adapté aux préadolescents, *Mesure et évaluation en éducation*, 31(1), 51-91.

Dierendonck, C., Milmeister, M., Milmeister, P., Weis, C., Fischbach, A., Ugen, S., & Martin, R. (2016, janvier). L'implication des élèves lors d'évaluations externes à faibles enjeux: Le cas de l'évaluation « Épreuves Standardisées » au Luxembourg. Contribution au 28ème colloque de l'ADMEE-Europe « Evaluations et apprentissages » à Lisbonne. 13-15 janvier 2016.

Dierendonck, C., Sonnleitner, P., Ugen, S., Keller, U., Fischbach, A., & Martin, R. (2013). La mesure de la motivation et de l'effort des élèves dans le cadre des Épreuves Standardisées au Luxembourg. Actes du congrès AREF 2013.

Eccles J. S., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, J.L., & Midgley, C. (1983). Expectancies, values, and academic bahaviors. In J.T. Spence (Ed.), *Achievement and acievement motivation* (pp. 75-146). San Francisco, CA: W.H. Freeman.

Eccles, J., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109-132.

Fishbein, M.A. et Ajzen, I. (1975). Belief, attitude, intention and behavior: an introduction to theory and research, Reading, MA, Addison Wesley.

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education, 16* (4), 277-298.

Keskpaik, S., & Rocher, T. (2012). Les évaluations à faibles enjeux : quel rôle joue la motivation? Une expérience à partir de PISA. Communication dans le cadre du 24e colloque de l'Admée-Europe, Luxembourg.

Keskpaik, S., & Rocher, T. (2013). La motivation des élèves français face à des évaluations à faibles enjeux : Comment la mesurer ? Quel impact sur les réponses ? Actes du congrès AREF 2013.

Marsh, H.W. (1988). Self-Description Questionnaire: A theoretical and empirical basis for the measurement of multiple dimensions of preadolescent self-concept. A test manual and a research monograph. San Antonio, TX: Psychological Corporation.

Meijer, R.R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement, 25* (2), 107-135.

Ndinga, P. & Frenette, E. (2010). Élaboration et validation de l'Échelle de motivation à bien réussir un test. *Revue Mesure et Évaluation en Éducation, 33*(3), 99-123.

Rios, J.A., Liu O.L., & Bridgeman, B. (2014). Identifying low-effort examinees on student learning outcomes assessment: A comparison of two approaches. *New Directions for Institutional Research, 161*, 69-82. En ligne 10.1002/ir.20068

Sundre, D. L. & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assessment Update*, 14 (1), 8-9.

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. Educational Pschology Review, 6, 49-78.

Wigfield, A., & Eccles, J. (1992). The development of achievement task values: A theoretical analysis. Developmental Review, 12, 265-310. Eklöf, H. (2006). Development and validation of scores from an instrument measuring student test-taking motivation. *Educational and Psychological Measurement*, 66, 643–656.

Wise, S. L., & DeMars, C. E. (2006). An application of item response time: The effort-moderated IRT model. *Journal of Educational Measurement, 43* (1), 19-38.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18* (2), 163-183. En ligne 10.1207/s15324818ame1802_2

**Annexe : Instrument de mesure de l'investissement des élèves lors d'un test**

| |
|---|
| Motivation générale pour l'école (3 items) |
| Dans l'ensemble, les activités scolaires m'intéressent. |
| La plupart du temps, j'aime aller à l'école. |
| Pour l'école en général, je suis très démotivé/très motivé. |
| Self-concept en mathématiques (3 items) |
| Les mathématiques m'intéressent. |
| J'aime les mathématiques. |
| Je suis bon(ne) en mathématiques. |
| Valeur projetée du test (4 items) |
| Ce test est utile. |
| Ce test est important. |
| Ce test est une perte de temps pour moi. |
| Je pense que la plupart des questions de ce test seront pas du tout intéressantes/très intéressantes. |
| Attente de réussite au test (3 items) |
| Je me sens capable d'obtenir un bon résultat à ce test. |
| Pour faire ce test, j'ai confiance en mes compétences. |
| Mon résultat à ce test sera probablement très mauvais/très bon. |
| Disposition à s'investir (6 items) |
| Je veux obtenir un bon résultat à ce test. |
| Pour ce test, j'ai l'intention de me concentrer. |
| Pour ce test, j'ai l'intention de travailler pas du tout sérieusement/tout à fait sérieusement. |
| Pour ce test, j'ai l'intention de faire un effort très faible/très élevé. |
| Ma motivation à atteindre un bon résultat lors de ce test est très faible/très élevée. |
| Pour ce test, je suis pas du tout motivé/tout à fait motivé. |
| Valeur perçue du test (4 items) |
| J'ai trouvé la plupart des questions de ce test intéressantes. |
| Faire ce test était utile. |
| Au final, ce test était une perte de temps pour moi. |
| Faire ce test était important. |
| Difficulté perçue du test (3 items) |
| C'était un défi de répondre correctement à la plupart des questions de ce test. |
| J'ai éprouvé des difficultés à répondre à la plupart des questions de ce test. |
| Dans l'ensemble, j'ai trouvé que ce test était très facile/très difficile. |
| Investissement effectif (6 items) |
| Pour ce test, j'ai voulu obtenir un bon résultat. |
| Pour ce test, j'ai été concentré(e). |
| Pendant ce test, j'ai été pas du tout motivé/tout à fait motivé. |
| Ma motivation à atteindre un bon résultat lors de ce test a été très faible/très élevée. |
| Pour ce test, j'ai fait un effort très faible/très élevé. |
| Pour ce test, j'ai travaillé pas du tout sérieusement/tout à fait sérieusement. |
| Investissement si le test avait compté pour des points (6 items) |
| Si ce test avait compté pour des points, j'aurais voulu obtenir un meilleur résultat. |
| Si ce test avait compté pour des points, j'aurais été davantage concentré(e). |
| Si ce test avait compté pour des points, j'aurais été pas du tout motivé/tout à fait motivé. |
| Si ce test avait compté pour des points, ma motivation à atteindre un bon résultat aurait été très faible/très élevée. |
| Si ce test avait compté pour des points, j'aurais fourni un effort très faible/très élevé. |
| Si ce test avait compté pour des points, j'aurais travaillé pas du tout sérieusement/tout à fait |

*Mots clés : évaluations externes, épreuves standardisées, motivation, effort, validité*

## *Analyse des tensions entre l'évaluation interne et l'évaluation externe en mathématiques, lecture et écriture en 6e année du primaire et de l'équivalence des épreuves externes*

**Micheline Joanne Durand, Marie-Hélène Asselin**

L'évaluation des apprentissages représente un aspect crucial du rôle de l'enseignant. Celui-ci, confronté à un nouveau référentiel de la démarche d'évaluation, doit porter un jugement professionnel respectant les valeurs sur lesquelles s'appuie le système éducatif du Québec. De plus, l'approche par compétences préconisée dans le Programme de formation de l'école québécoise et la Politique d'évaluation des apprentissages (MEQ, 2003) viennent encore complexifier cette démarche évaluative. Les enseignants ont à planifier et à évaluer les apprentissages réalisés, c'est-à-dire à offrir des situations qui permettent d'inférer la compétence développée et à porter un jugement professionnel sur la qualité de la compréhension des élèves tant en cours qu'en fin de cycle (Durand et Chouinard, 2012). Ainsi, les élèves sont amenés à formuler des réponses élaborées pour résoudre des situations complexes qui exigent la mobilisation d'un ensemble de ressources. À cet effet, le ministère indique dans son référentiel de compétences professionnelles que les enseignants doivent être aptes à « Évaluer des apprentissages et le degré d'acquisition des compétences des élèves pour les contenus à faire apprendre.(compétence 5)» Il soumet également des épreuves en français et en mathématiques pour l'évaluation des élèves de 6e année qui servent de modèles de situations complexes pour les enseignants.

Le présent projet de recherche porte sur le jugement professionnel des enseignants, l'évaluation des compétences et les pratiques évaluatives. Il vise l'amélioration de la persévérance et de la réussite scolaires par l'examen des pratiques évaluatives qui sous-tendent le jugement professionnel porté par les enseignants en cours et en fin de cycle. Plus spécifiquement, il analyse l'approche évaluative adoptée par les enseignants de la 6e année du primaire à toutes les étapes de la démarche évaluative (planification, collecte, interprétation et jugement) et sa concordance avec la réussite des élèves aux épreuves ministérielles.

L'évaluation des compétences semble provoquer l'émergence d'une multitude d'approches évaluatives chez les enseignants. Plusieurs utiliseraient même des pratiques évaluatives peu conformes à ce qui est attendu (McMillan, 2001). Cela dit, les recherches disponibles actuellement renseignent relativement peu sur les connaissances et les pratiques des enseignants du primaire en ce qui concerne l'évaluation des apprentissages dans un contexte d'approche par compétences (Brookhart, 1993, 2004 ; Goodman et Hambleton, 2004 ; McMillan, 2000, 2001).

Notre projet vise à répondre principalement aux deux questions suivantes : comment les enseignants de la 6e année du primaire documentent-ils leur jugement (approches évaluatives et conditions d'application) à chaque étape de la démarche d'évaluation en cours et en fin de cycle en regard des compétences en écriture, lecture et mathématiques? À partir des pratiques d'évaluation formatives réputées efficaces dans la littérature, quels sont les différents profils qui se dégagent et sont-ils en lien avec les données produites par les épreuves ministérielles? Ces questions de recherche se traduisent par les objectifs suivants : 1) Recenser les pratiques mises en place par les enseignants pour évaluer les compétences en lecture, en écriture et en mathématiques de leur élèves, et ce, à chaque étape de la démarche d'évaluation en cours de cycle, 2) Comparer ces pratiques avec celles mises en place en fin de cycle, 3) Dégager des profils des pratiques d'évaluation formative des enseignants de la 6e année du primaire et évaluer l'importance proportionnelle de chacun selon l'âge, l'expérience et la formation des enseignants, 4) Identifier les tensions entre les résultats des élèves aux épreuves internes (bulletin et bilan) et les données obtenues à l'aide des épreuves ministérielles en tenant compte de la discipline concernée. Ces objectifs veulent favoriser le développement de connaissances dont le but est de soutenir adéquatement les pratiques d'évaluation des enseignants.

En juin 2010 et 2011, les enseignants de 6e année du primaire ont fait exécuter à leurs élèves un modèle national d'évaluation (épreuves uniques). Ce sont les résultats obtenus à ces épreuves, ceux obtenus au cours des deux années du cycle (notes au bulletin) et le jugement porté par les enseignantes au bilan de fin de cycle qui forment la base des données de notre étude quant aux résultats des élèves.

L'analyse des résultats a permis de faire un portrait global de la manière dont l'évaluation est faite par 55 enseignants de 6e année. À la lumière de ces analyses, on peut constater que la formation continue n'est pas offerte de manière égale à tous les enseignants : ceux du privé semblent y avoir moins accès que ceux travaillant dans le système public. De plus, on constate que les formations permettant un réinvestissement concret dans la pratique des enseignants ont été davantage suivies que celles traitant des fondements et des orientations de l'évaluation. Toutefois, on remarque que la formation, principalement offerte par les commissions scolaires, n'apporte pas nécessairement un changement de pratiques chez les enseignants. Est-ce à cause des conditions dans lesquelles se tiennent les formations, de leur qualité, ou est-ce plutôt un refus de la part des enseignants de changer leurs pratiques?

L'analyse des résultats a également montré que les enseignants n'accordent pas tous de l'importance à baser leur pratique sur les documents ministériels : plusieurs enseignants agissent selon ce qu'ils pensent adéquat de faire. On remarque également que malgré le changement d'orientation du Renouveau pédagogique de 2002, les enseignants n'endossent pas complètement les nouvelles perspectives d'évaluation s'y rattachant. Ils alternent selon les contextes entre une évaluation plus traditionnelles et des perspectives nouvelles. L'intégration n'est pas complétée et on perçoit encore une certaine confusion chez plusieurs enseignants quant aux critères et aux outils d'évaluation à utiliser ou encore aux aspects à prendre en compte pour attribuer une note au bulletin. On dénote aussi une sous-utilisation des mesures de différenciation pour les élèves en difficulté et quand elles sont mises en place, elles touchent surtout les structures de travail ou les stratégies et processus. Pour favoriser la réussite des élèves, les enseignants pourraient faire appel à une plus grande variété de mesures qui répondraient réellement aux besoins de l'élève.

Face aux disparités perçues dans les pratiques évaluatives des enseignants sondés, on peut certainement conclure qu'il faudrait offrir de la formation pour aider les enseignants québécois à endosser les orientations du Programme de formation et les pratiques prescrites par la Politique en évaluation des apprentissages. Puisque le changement est un processus complexe, il faudrait penser à des formations qui durent dans le temps : l'accompagnement par un professionnel, l'adhésion à une communauté d'apprentissage professionnelle ou encore la participation à un programme universitaire en évaluation des compétences par exemple.


BROOKHART, S.M. (1993). Teacher's grading practices : Meaning and values. *Journal of Educational Measurement,* 30(2), 123-142.

BROOKHART, S.M. (2004). Grading. Upper Saddle River, NJ : Pearson, Merrill, Prentice-Hall.

DURAND. M.J. et CHOUINARD. R. (s. dir.) (2012*) L'Évaluation des apprentissages, de la planification de la démarche à la communication des résultats*. Montréal : Édition revue et augmentée, Édition Marcel Didier.

GOODMAN, D. P., & HAMBLETON, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. Applied Measurement in Education, 17, 145–220.

MCMILLAN, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, 7(8).

MCMILLAN, J. H. (2001). *Essential assessment concepts for teachers and administrators*. Thousand Oaks, CA: Corwin Publishing Company.

MINISTÈRE DE L'ÉDUCATION DU QUEBEC (2003) *Politique d'évaluation des apprentissages ; formation générale des jeunes, formation générale des adultes et formation professionnelle.* Québec : Gouvernement du Québec.



*Mots clés : épreuves externes, pratiques évaluatives, enseignement primaire*