

Mesure et évaluation en éducation



L'évaluation de la qualité de dispositifs scolaires Développement d'une approche multidimensionnelle et bidirectionnelle

Philippe Tremblay

Volume 35, numéro 2, 2012

Date de réception: 31 mars 2011

Date de réception de la version finale : 28 mai 2012

Date d'acceptation: 18 juin 2012

URI : <https://id.erudit.org/iderudit/1024718ar>

DOI : <https://doi.org/10.7202/1024718ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

ADMEE-Canada - Université Laval

ISSN

0823-3993 (imprimé)

2368-2000 (numérique)

[Découvrir la revue](#)

Citer cet article

Tremblay, P. (2012). L'évaluation de la qualité de dispositifs scolaires : développement d'une approche multidimensionnelle et bidirectionnelle. *Mesure et évaluation en éducation*, 35(2), 39–68.
<https://doi.org/10.7202/1024718ar>

Résumé de l'article

Dans le cadre d'une recherche portant sur l'évaluation comparée de la qualité de deux dispositifs scolaires destinés à des élèves ayant des troubles d'apprentissage, l'auteur a été amené à développer, sur la base de concertations avec différents groupes d'intervenants, de concepts et de théories émanant des sciences de l'éducation, un modèle multidimensionnel et bidirectionnel de la qualité des dispositifs scolaires. Délaissant le « débat ontologique » (Behrens, 2007) autour du concept de qualité, ce travail de conceptualisation et de formalisation d'un cadre d'évaluation de la qualité de dispositifs s'est inspiré de travaux similaires (Bou - chard & Plante, 2002 ; De Ketele & Gerard, 2007) portant sur des objets proches (programme, organisme, etc.). Sur la base des liens de conformité entre un référent et un référé, issus des caractéristiques d'un dispositif, l'auteur a construit un modèle composé de dix dimensions bidirectionnelles (pertinence, adéquation, congruité, synergie, efficacité, efficience, fiabilité, impact, bien-fondé et flexibilité).

L'évaluation de la qualité de dispositifs scolaires : développement d'une approche multidimensionnelle et bidirectionnelle

Philippe Tremblay

Université Laval

MOTS CLÉS : évaluation, qualité, dispositif, programme, modèle théorique

Dans le cadre d'une recherche portant sur l'évaluation comparée de la qualité de deux dispositifs scolaires destinés à des élèves ayant des troubles d'apprentissage, l'auteur a été amené à développer, sur la base de concertations avec différents groupes d'intervenants, de concepts et de théories émanant des sciences de l'éducation, un modèle multidimensionnel et bidirectionnel de la qualité des dispositifs scolaires. Délaissant le « débat ontologique » (Behrens, 2007) autour du concept de qualité, ce travail de conceptualisation et de formalisation d'un cadre d'évaluation de la qualité de dispositifs s'est inspiré de travaux similaires (Bouchard & Plante, 2002 ; De Ketele & Gerard, 2007) portant sur des objets proches (programme, organisme, etc.). Sur la base des liens de conformité entre un référent et un référé, issus des caractéristiques d'un dispositif, l'auteur a construit un modèle composé de dix dimensions bidirectionnelles (pertinence, adéquation, congruité, synergie, efficacité, efficience, fiabilité, impact, bien-fondé et flexibilité).

KEY WORDS: evaluation, quality, model, programme, theoretical model

In a study comparing the level of quality of two education programmes for students with learning difficulties, and through interviews with various groups of educators and a review of concepts and theories in education sciences, the author developed a multidimensional and bidirectional model to evaluate this aspect of education programmes. Away from the ontological debate (Behrens, 2007) over the notion of quality, our conceptualization and formalization of this evaluatory method was inspired by similar research on the subject (Bouchard & Plante, 2002 ; De Ketele & Gerard, 2007) with subject proximity (programme, institution). Based on the compliance between referent and referred stemming from various characteristics of an education programme, the resulting approach consists of ten bidirectional dimensions (relevance, agreement, congruity, synergy, effectiveness, efficiency, reliability, impact, validity, and flexibility).

PALAVRAS-CHAVE: avaliação, qualidade, modelo programa, modelo teórico

No quadro de uma investigação sobre a avaliação comparada da qualidade de dois dispositivos escolares destinados a alunos com problemas de aprendizagem, o autor desenvolveu, a partir do diálogo com diferentes grupos de intervenientes e da revisão de conceitos e de teorias das ciências da educação, um modelo multi-dimensional e bidirecional da qualidade dos dispositivos escolares. Afastando-se do “debate ontológico” (Behrens, 2007) em torno do conceito de qualidade, este trabalho de conceptualização e de formalização de um quadro de avaliação da qualidade dos dispositivos inspirou-se em trabalhos similares (Bouchard & Plante, 2002 ; De Ketele & Gerard, 2007) sobre objetos próximos (programa, organismo, etc.). Baseado nos laços de conformidade entre um referente e um referido, resultantes das características de um dispositivo, o autor construiu um modelo composto de dez dimensões bidirecionais(pertinência, adequação, congruência, sinergia, eficácia, eficiência, fiabilidade, impacto, validade e flexibilidade).

Note de l’auteur – Toute correspondance peut être adressée comme suit: Philippe Tremblay, professeur, Université Laval, Faculté des sciences de l’éducation, bureau 1034, 2320, rue des Bibliothèques, Québec (QC) G1V 0A6, téléphone : (418) 656-2131 poste 6151, télécopieur: (418) 656-2905, ou par courriel à l’adresse suivante : [philippe.tremblay@fse.ulaval.ca].

Introduction

Le concept de qualité a pris une importance notable ces dernières années en sciences de l'éducation (Beamish, 2004 ; Behrens, 2007 ; Bouchard & Plante, 2002 ; UNESCO, 2005 ; Winch, 1996). Toutefois, ce concept est polysémique, multidimensionnel, à plusieurs niveaux, et dynamique (Vlăsceanu, Grünberg, & Pârlea, 2007). Il est multifacette, contesté et jamais pleinement représentable (Stake & Schwandt, 2006).

À titre d'exemple, Adams (1993) recense une cinquantaine de définitions de ce terme. Harvey (1995) classe les définitions de la qualité en cinq catégories :

- 1) exception ou excellence (Arcaro, 1995 ; Doherty-Delorme & Shaker, 2001),
- 2) la perfection/ou l'absence de défaut (Poole, 2010 ; Shaw, 2010 ; Wangenge-Ouma & Langa, 2010),
- 3) l'atteinte de standards, de repères et d'indicateurs (Beamish, 2004 ; Otis-Wilborn, Winn, Ford, & Keyes, 2000),
- 4) l'ajustement aux besoins des consommateurs (*Council for Higher Education Accreditation*, 2002 ; Juran, 1993 ; Tempus, 2001),
- 5) une valeur monétaire (Winch, 1996) ou une amélioration/transformation (Campbell & Rozsnyai, 2002 ; Van Kemenade, Pupius, & Hardjono, 2008).

La qualité est utilisée tant pour décrire que pour juger ou apprécier un objet (Stake & Schwandt, 2006). D'une part, il est question des qualités d'un objet en se référant aux différentes propriétés qui le constituent ou le caractérisent. La qualité devient synonyme de spécifications. D'autre part, la qualité est aussi communément considérée comme un jugement ou une évaluation de l'excellence, du mérite ou de la valeur sur la base de différents critères (objectifs, résultats, etc.). La distinction faite par Torres (1996) entre deux conceptions de la qualité : spécifications d'un objet (*héxis*) versus un acte de jugement (*poïos*) rejoint cette première distinction. L'*héxis* est défini comme la conformité « à l'ensemble des caractéristiques individuelles qui la constituent et l'identifient comme ce qu'elle est » (Torres, 1996, p. 34). Dans ce cadre, la qualité s'identifie, d'une part, à l'excellence (ou la perfection) et, d'autre part, à la différenciation. Le *poïos*, quant à lui, est « un acte de jugement exprimé

à partir de normes et de critères préétablis relativement à un objet quelconque» (Torres, 1996, p. 35). Dans ce second cas, un dispositif est de qualité lorsqu'il reprend une série de normes, de standards, et qu'il a été jugé (évalué) conforme à ceux-ci. Dans ce cas, la qualité est également vue comme mesurable et précise.

L'évaluation de la qualité des programmes et des dispositifs'

Le but principal de l'évaluation de programme est de déterminer la qualité de ce dernier en formulant un jugement (Hurteau, Valois, & Bossiroy, 2011 ; Stake & Schwandt, 2006). Stake et Schwandt (2006) définissent deux catégories pour conceptualiser la qualité et appuyer ce jugement: la qualité mesurée et la qualité ressentie. Pour la première, la qualité est considérée comme mesurable et le jugement se fait à partir de différents critères. Ce jugement à partir de critères crée une distance où les concepts et le langage des sciences sociales sont utilisés. La qualité est structurée dans une théorie de l'évaluation à partir d'une série de construits. La valeur de ces modèles d'évaluation dépend de l'ingénuité de l'évaluateur à représenter ces construits avec validité et parcimonie (Stake & Schwandt, 2006). Pour la seconde catégorie, la qualité est dépendante de l'expérience personnelle de différents intervenants directement en interaction avec le dispositif. L'emphase est mise tant sur le sens subjectif et intersubjectif donné à l'expérience. Le savoir pratique développé devient la source du discernement de la qualité. La proximité de l'évaluateur est essentielle pour saisir le sens de l'expérience.

Fitzpatrick, Sanders, et Worthen (2011), quant à eux, ont dégagé quatre approches évaluatives selon quatre catégories basées sur les facteurs primaires guidant ou dirigeant l'évaluation: le jugement éclairant la qualité d'un programme (par ex. : expertise, consommateur), l'amélioration du programme (par ex. : objectifs, standards, théorie), les décisions devant être envisagées (Patton, 2008; Stuffelbeam & Shinkfield, 2007) et la participation des intervenants (par ex. : évaluation développementale, autonomisation, démocratisation).

Les modèles logiques

Depuis plus d'une trentaine d'années, les évaluateurs en sciences sociales ou en sciences de l'éducation ont développé nombre d'approches basées sur une théorie du programme, d'une logique du programme ou encore d'un modèle logique. Porteous (2009) définit un programme comme: «une série d'activités appuyées par un ensemble de ressources et visant à atteindre des résultats spécifiques au sein de groupes cibles définis» (p. 87). Un programme ou un dispositif apparaît comme un outil de transformation (Riddle & Dagenais, 2009).

Ce qui unit ces approches est la reconstruction d'un modèle causal basé sur diverses sources pour comprendre comment un dispositif ou un programme spécifique est censé produire des extrants, des résultats (Hansen & Vedung, 2010; Stame, 2004). Selon Chen (2005), le modèle est un ensemble d'hypothèses implicites et explicites expliquant comment le programme devrait être organisé et pourquoi on s'attend à ce qu'il fonctionne. Pour Porteous (2009), en l'absence d'un tel outil, l'élaboration d'un modèle logique devient la première étape de tout processus d'évaluation.

Chen (1990) fait la distinction entre deux méthodes de développement d'un modèle théorique d'évaluation d'un dispositif ou d'un programme :

- a) l'approche par intervenants (ou détenteurs d'enjeux) où l'évaluateur travaille avec ces derniers, principalement des acteurs clés, pour découvrir leur raisonnement et les principes sous-jacents affectant le succès du dispositif;
- b) l'approche par les sciences sociales où l'évaluateur fait usage tant de sa connaissance du dispositif que des concepts et des recherches en sciences sociales, pour développer un modèle théorique.

Divers auteurs insistent sur l'importance d'intégrer les intervenants dans l'élaboration du modèle théorique (Fitzpatrick et al., 2011 ; Rossi, Freeman, & Lipsey, 1999).

Selon Birckmayer et Weiss (2000), les sources possibles pour élaborer un modèle théorique sont les sciences sociales elles-mêmes et la recherche, les évaluations antérieures, les attentes des gestionnaires et praticiens, les connaissances de l'évaluateur et son expérience avec des dispositifs du même type et, enfin, sa pensée logique. Le plus souvent, l'évaluateur fera des cycles entre ces différentes sources puis négociera avec les intervenants pour arriver à un commun accord. On tente de créer un modèle unique intégrant les conceptions

de l'ensemble des intervenants dans une conception unique derrière laquelle chacun doit se rallier (Hansen & Vedung, 2010). Donaldson (2007), quant à lui, décrit quatre sources potentielles pour l'élaboration d'un modèle théorique :

- 1) une théorisation précédente et les travaux de recherche antérieurs,
- 2) des théories implicites proches du programme,
- 3) l'observation du programme en opération et,
- 4) la recherche exploratoire pour tester des hypothèses critiques relatives à un présumé modèle théorique.

Les questions posées dans le cadre d'une évaluation de programme peuvent provenir de la théorie, mais également d'une conceptualisation à partir des caractéristiques de l'objet évalué (Fitzpatrick et al., 2011), de ses objectifs, de son environnement et de sa population, des intervenants de différents niveaux et des besoins des évaluateurs (Levin-Rozalis, 2009).

Un aspect crucial de la théorie du programme est comment ses différentes composantes sont liées les unes aux autres (Davidson, 2000, 2005). Pour décrire un programme, il faut donc fournir une compréhension de la manière dont différents événements, personnes, fonctions et autres éléments représentés dans la théorie sont supposément liés (Rossi et al., 1999). Il est important de pouvoir décrire le dispositif, ses composantes, et ce, en gagnant l'accord de différents groupes d'intervenants. «La description précise d'un programme et des liens entre ses composantes est indispensable à tout processus évaluatif» (Riddle & Dagenais, 2009, p. 21). Riddle et Dagenais (2009) formalisent huit composantes d'un programme: besoins, objectifs, intrants, processus et extrants, participants, effets à court et moyen termes, effets à long terme et facteurs externes sur les effets du programme. Porteous (2009), quant à lui, définit trois éléments essentiels d'un programme (composantes et activités, groupes cibles, résultats) et quatre éléments facultatifs: intrants, extrants, facteurs contributifs, facteurs externes.

Les éléments utilisés pour décrire ou représenter une théorie de programme incluent ainsi, le plus souvent, des intrants (par ex.: ressources, population, etc.), des activités (actions et rétroactions) et des extrants (effets de l'action) qui sont mis en relation avec des résultats initiaux, intermédiaires et de long terme (Coryn, Noakes, Westine, & Schröter, 2011; Donaldson, 2007; Lipsey, Rossi, & Freeman, 2004; Patton, 2008). Des différentes composantes d'un programme ou dispositif dans la documentation, l'auteur distingue trois niveaux: composantes de base, actions et effets. Les composantes de base

constituent le « corps » du dispositif, la partie définie, visible dans un environnement, un contexte. Ces composantes, ayant contraintes et possibilités, agissent en poursuivant tant la mission prioritaire du dispositif (actions) que celles visant sa pérennisation/amélioration (rétroaction). Enfin, ces actions produisent des traces, des effets dans des temps et des espaces variables. Ces effets affectent les mêmes composantes de base. Elles impliquent que le dispositif ou le programme agit dans un espace/temps, c'est-à-dire qu'il connaît un vieillissement et une déformation.

La plupart des approches basées sur la théorie partagent trois caractéristiques fondamentales :

- 1) expliquer la théorie d'un programme en détaillant les relations supposées entre les intrants, les processus, médiateurs et les résultats à court et long termes,
- 2) analyser tous les construits postulés dans la théorie et,
- 3) traiter les données pour vérifier comment les relations postulées apparaissent en réalité (Shadish, Cook, & Campbell, 2002).

Enfin, le modèle doit être plausible (avoir une apparence de vérité, de raison et de crédibilité) et stipuler les séquences de cause à effet à travers lesquelles les actions sont présumées produire des effets à long terme (Donaldson & Lipsey, 2006; Stame, 2004). L'évaluation sur la base d'un modèle théorique a deux composantes vitales : le premier est conceptuel tandis que le second est empirique (Coryn et al., 2011; Rogers, Petrosino, Huebner, & Hacsí, 2000). Cette évaluation doit tant expliquer le modèle qu'investiguer les causes et les effets observés.

Dans le cadre d'une recherche portant sur l'évaluation comparée de la qualité de deux dispositifs scolaires destinés à des élèves ayant des troubles d'apprentissage, l'auteur a été amené à développer, sur la base de concertations avec différents groupes d'intervenants (macros, mezzos et micros) et de concepts et théories émanant des sciences de l'éducation, un modèle multidimensionnel et bidirectionnel de la qualité des dispositifs scolaires. Délaissant le « débat ontologique » (Behrens, 2007) autour du concept de qualité, ce travail de conceptualisation et de formalisation d'un cadre d'évaluation de la qualité de dispositifs s'est inspiré de travaux similaires (Bouchard & Plante, 2002; De Ketele & Gerard, 2007) portant sur des objets proches (programme, organisme, etc.).

Contexte entourant le développement du modèle

En Communauté française de Belgique, un enseignement spécialisé, distinct de celui de l'enseignement ordinaire, est organisé aux ordres d'enseignement maternel, primaire et secondaire pour les élèves à besoins spécifiques. Il est composé de huit types d'enseignement, correspondant à différentes catégories de déficiences et de troubles (physiques, sensorielles, intellectuelles, troubles d'apprentissage, etc.).

À la suite de travaux antérieurs portant sur son évaluation (Communauté française de Belgique, 2007 ; Delvaux, 2000 ; Franssen & Van Campenhoudt, 2004 ; Tremblay, 2007, 2011), différents constats, posant de sensibles questions éthiques (catégorisation, faible espérance scolaire, stigmatisation, etc.), budgétaires (accroissement de la population, maintien dans les structures spécialisées à long terme, etc.), pédagogiques (difficulté de l'enseignement spécialisé et ordinaire à réintégrer ces élèves, gestion de l'hétérogénéité, etc.) et politiques (pression des parents, pression parlementaire, etc.), ont amené chercheurs et acteurs institutionnels à amorcer une réflexion sur des dispositifs pédagogiques originaux permettant de répondre aux difficultés éprouvées par l'enseignement spécialisé.

Dispositifs et clientèle

Sur la base de contraintes budgétaires et légales, voulant que la population prise en compte (troubles d'apprentissage) et les ressources déployées soient les mêmes qu'en enseignement spécialisé, un dispositif novateur, basé sur l'inclusion groupée, en classe régulière, d'environ six élèves ayant des troubles d'apprentissage et d'un coenseignement, à temps plein, entre un enseignant ordinaire et un orthopédagogue, a été mis en place. Ces expériences débutaient en première année du primaire dans une classe ordinaire (de 12 à 22 élèves) où étaient intégrés trois à sept élèves ayant des troubles d'apprentissage². Ce dispositif de coenseignement visait à être comparé à des classes d'enseignement spécialisé « classique ». Ces classes de l'enseignement spécialisé étaient du même ordre d'enseignement scolaire et également destinées à des élèves présentant des troubles d'apprentissage. En enseignement spécialisé, un instituteur enseigne à un groupe d'élèves plus réduit (de 3 à 12 élèves par classe) qui reçoivent également des interventions de logopédie/orthophonie. Notre échantillon était composé de huit classes inclusives et de neuf classes d'enseignement spécialisé de 1^{re} et de 2^e années de l'enseignement primaire.

Processus de concertation avec les intervenants

L'auteur a utilisé la comparaison pour faciliter le discernement de la qualité entre les deux dispositifs et a effectué un choix des dimensions et des critères en relation avec les intervenants (Stake & Schwandt, 2006). L'auteur s'est inspiré du processus explicité par Donaldson (2007) et par Birckmayer et Weiss (2000). Ceux-ci proposent différents allers/retours entre terrain et théorie, allant de plus en plus en profondeur dans l'analyse du dispositif et dans la pertinence du modèle théorique. Dans le présent cas, chaque aller/retour a été marqué par les apports d'un nouveau groupe d'intervenants. Ces groupes d'intervenants se rapprochent de ceux définis par Greene (2006) : administrateurs d'un programme, gestionnaires, clients et ceux qui sont directement affectés.

Dans la situation présente, le passage de dispositifs d'enseignement spécialisé ségrégatifs vers d'autres, plus intégratifs ou inclusifs, s'est étroitement accompagné de recherches empiriques cherchant à démontrer l'efficacité relative de l'intégration/inclusion par rapport à la ségrégation et s'est appuyé sur celles-ci. Cette évaluation comparative, dans ce contexte précis, répondait au postulat « originel » de l'enseignement spécialisé ségrégatif qui avançait que son organisation spécifique, en dehors de l'enseignement ordinaire, était plus efficace pour prendre en charge ces élèves (Vienneau, 2002). Dans ce cadre, l'efficacité des dispositifs, en termes de performances scolaires, psychosociales ou comportementales, était souvent confondue avec leur qualité. Cela peut expliquer pourquoi, dans les premières phases de nos travaux, les chercheurs et les intervenants institutionnels (ou « macros ») eurent principalement comme priorité de démontrer de manière comparative l'efficacité du dispositif inclusif « nouveau » par rapport à un « ancien », mais également de vérifier si le dispositif inclusif ne nuisait pas aux élèves sans difficulté y participant (impact). D'autres considérations concernant l'efficience et l'équité furent également mises de l'avant.

Lors de la mise en œuvre du dispositif, de nouveaux intervenants intermédiaires ou « mezzo » (directions d'école, responsables de Pouvoir Organisateur, centres orienteurs) se sont joints aux entretiens et groupes de discussion (Krueger & Casey, 2000; Morgan, 1997), offrant une confrontation avec les premières dimensions choisies pour l'évaluation du dispositif. Enfin, lors de l'implémentation proprement dite des projets pilotes, des intervenants du terrain ou « micros » se sont encore greffés à notre réflexion (enseignants et orthophonistes). Chacun des groupes d'intervenants est venu avec des

demandes spécifiques concernant l'évaluation des dispositifs. Tel qu'indiqué précédemment, les intervenants institutionnels ou macros ($n = 6$) étaient préoccupés principalement, en ce qui concerne l'évaluation des dispositifs, par l'efficacité, l'efficience et l'équité. Les rencontres avec les intervenants mezzo ($n = 27$) ont montré que ceux-ci étaient principalement préoccupés par la validité du recrutement des élèves, des possibles effets collatéraux chez les élèves sans difficulté, de l'attitude des parents, de l'organisation des ressources, des possibilités de flexibilité du dispositif. Enfin, les intervenants micro ($n = 23$) ont insisté, notamment, sur l'importance de leur perception face à la population et de l'efficacité, aux attitudes des élèves, des enseignants, parents et élèves. Par ailleurs, de manière transversale, au regard des résistances envers l'inclusion scolaire chez certains acteurs, des questionnements plus larges sur les objectifs, aux besoins, aux ressources ou aux effets ont été émis (par ex. : Ces objectifs sont-ils atteignables pour cette population ? Avec ces ressources ? Ces effets sont-ils souhaitables ? Lesquels ? Remplir l'objectif de réussite scolaire est-il bénéfique ou comporte-t-il des risques ?). Enfin, au regard de la diversité des classes/écoles participantes et de la nouveauté du dispositif inclusif, l'évaluation des variations dans le temps et l'espace des dispositifs étudiés est ensuite apparue comme une priorité. Cet éclatement de dimensions à prendre en compte dans l'évaluation de dispositif a mené à rechercher des modèles permettant leur intégration.

L'absence d'utilisation de modèles existants, la taille relativement réduite des expériences et la relative proximité des deux dispositifs ont mené à chercher une théorie utilisable pour l'ensemble des intervenants. Par ailleurs, l'utilisation d'une approche comparative a incité à rechercher un modèle flexible, pouvant être utilisé tant pour l'un que pour l'autre dispositif. De plus, il fallait, pour éviter certaines résistances et refus de collaborer, disposer d'un modèle suffisamment précis dans les dimensions à évaluer et leur justification, mais suffisamment large et neutre pour être accepté de tous. Enfin, ce type d'évaluation requerrait d'articuler des méthodes mixtes, qualitatives et quantitatives (Creswell, 2009; Pluye et al., 2009), au regard de dispositifs évolutifs situés et de perspectives d'intervenants plurielles.

La concertation avec les différents niveaux d'intervenants (macro, mezzo et micro) a conduit, dans un premier temps, à lister les différents indicateurs issus de leurs préoccupations puis à les catégoriser selon les premières dimensions ayant émergé des discussions (efficacité, pertinence, etc.). Par la suite, un consensus, pour chaque niveau d'intervenants, a été recherché. Des modèles théoriques d'évaluation de la qualité de dispositifs ont alors été utilisés

pour tenter de lier ces dimensions et donner des fondements conceptuels aux dimensions émergeant de la concertation avec les intervenants. L'auteur a utilisé deux modèles d'évaluation de la qualité: celui de Bouchard et Plante (2002) puis celui de De Ketele et Gerard (2007).

Apports et limites des deux modèles

Le premier modèle utilisé, celui de Bouchard et Plante (2002), distingue trois types de qualité: les qualités singulières, la qualité postulée et les qualités transversales. Les qualités singulières sont des attributs que possèdent, en propre et pour eux-mêmes, divers constituants d'un organisme. «La clarté des objectifs et la clarté des besoins par exemple constituent des qualités singulières» (Bouchard & Plante, 2002, p. 226). La qualité postulée fait appel à ce qui est perçu par les usagers. Enfin, les qualités transversales portent sur le degré de conformité entre les six caractéristiques fondamentales d'un organisme, à savoir: des besoins, des contraintes, des objectifs, des moyens, des ressources, des effets ou des résultats. Cette qualité est relative et se comprend comme le rapport entre un «référé» (objet sur lequel on recueille une information) et un «référentiel» (critère de comparaison). Neuf qualités transversales sont identifiées: pertinence, à-propos, efficacité, efficience, impact, cohérence, synergie, durabilité et flexibilité. Ces qualités, inspirées de la documentation, s'articulent au sein d'un modèle circulaire, semi-fermé.

Le modèle de De Ketele et Gerard (2007) s'inspire du modèle de Roegiers (2003) et de Bouchard et Plante (2002). De Ketele et Gerard (2007) définissent 15 dimensions (pertinence, réalisme politique, validité, cohérence, faisabilité, applicabilité, régulation des moyens, efficacité interne, efficacité externe, durabilité, efficience, adhésion, synergie, conformité et équité) sur la base de lien de conformité entre trois environnements (normes, besoins et enjeux personnels des acteurs) et de trois composantes (objectifs/effets attendus, produits/impacts et ressources/stratégies). De Ketele et Gerard (2007) avancent que la qualité est toujours relative et s'énonce donc essentiellement en termes de rapports entre un objet sur lequel on recueille une information, le référé et un référentiel qui sert de critère à la comparaison. On remarque que différentes composantes du système éducatif ont des doubles rapports. C'est le cas, par exemple, entre les objectifs et les besoins (pertinence et réalisme politique), les résultats et les ressources/contraintes/stratégies (efficience et optimisation). Le modèle de De Ketele et Gerard (2007) est toutefois statique et la dimension temporelle y est peu présente.

Si on analyse ces deux tentatives de définition de la qualité des dispositifs en relation avec les apports des intervenants, différents constats sur les apports et les limites sont à mettre en évidence. Tout d'abord, dans ces deux modèles, le concept de qualité est étroitement associé au concept de « conformité » ou d'adéquation. Cette qualité est relative et se comprend comme le rapport entre un « référé » (objet sur lequel on recueille une information) et un « référent » (critère de comparaison). Les auteurs empruntent à Figari (1994) le processus de référentialisation. Cette conformité implique une mesure et une évaluation de ce lien de conformité entre les différentes caractéristiques de l'objet ; la qualité se situant dans les liens entre ces caractéristiques. Bien que les dimensions puissent varier d'un auteur à l'autre, un certain consensus se dégage sur différentes caractéristiques à prendre en compte lors de l'évaluation de la qualité d'un système scolaire (apprenants, contenus, processus, résultats, contextes, etc.), de dimensions qui composent la qualité (efficacité, pertinence, cohérence, etc.).

Ces définitions de la qualité ont pu servir de base de concertation permettant d'intégrer les différentes demandes des trois groupes d'intervenants. Toutefois, la représentation graphique du modèle présenté par Bouchard et Plante (2002) présente différentes difficultés limitant sa compréhension et sa cohérence selon plusieurs intervenants consultés. Plusieurs éléments reviennent à divers endroits du schéma, parfois sous des formes différentes. Ainsi, les besoins sont présents à deux endroits, tout comme l'environnement/contrainte. Les dimensions sont unidirectionnelles, semblant aller dans le sens de l'énoncé du lien de conformité. Le référent et le référé sont figés dans un rapport préétabli. Enfin, les effets d'impact ne sont dirigés que vers l'environnement et l'équité y est absente.

Chez De Ketele et Gerard (2007), plusieurs dimensions ont une dimension « miroir » interrogeant inversement le référé et le référent. C'est le cas, par exemple, de la pertinence (objectifs/référé vs besoins/référentiel) et du réalisme politique (objectifs/référentiel et besoins/référé). Cette caractéristique correspond aux demandes des intervenants, car il permet d'inverser explicitement le questionnement entre deux composantes du dispositif, de questionner les objectifs ou les ressources en rapport avec les objectifs ou la population. Cependant, de fait, on accumule les dimensions plutôt que de percevoir deux facettes d'une même dimension. Cette non-prise en compte de la bidirectionnalité des dimensions conduit à un modèle manquant de parcimonie, plus lourd à gérer ou incomplet, car certains rapports inversés entre compo-

santes sont passés sous silence. Le lien de conformité entre caractéristiques de l'objet évalué pourrait donc être compris comme un double rapport, c'est-à-dire une double relativité. En somme, le rapport de conformité définit la dimension, et non la direction de ce rapport (référentiel/référé).

Enfin, les effets du dispositif ne sont pas ou peu perçus comme pouvant générer des métaeffets. C'est ce que l'auteur pourrait appeler le « vieillissement » du dispositif. Dans le modèle de De Ketele et Gerard (2007), ces effets sont principalement centrés sur les besoins. Bouchard et Plante (2002), dans un modèle circulaire et flexible, induisent cette gestion des effets dans le temps de manière explicite. Cependant, ces effets sont orientés, là aussi, vers les besoins et vers l'environnement, mais pas sur les autres composantes du dispositif (objectifs et ressources). Dans le modèle de De Ketele et Gerard (2007), les effets d'impact n'apparaissent explicitement pas dans le modèle.

Proposition d'un modèle d'évaluation des dispositifs scolaires : les composantes du dispositif

Pour les apports théoriques et conceptuels, outre les modèles de Bouchard et Plante (2002) et de De Ketele et Gerard (2007), l'auteur s'est également basé sur les travaux de Figari (1994, 2001, 2006) qui propose une méthodologie de référentialisation pour accompagner la modélisation. Figari (1994) définit la référentialisation comme la recherche des éléments pouvant expliquer et justifier la programmation, l'organisation et l'évaluation des actions pédagogiques. L'évaluation est comprise comme une activité de comparaison entre un objet observé et un système de référence. Cette conformité implique une mesure et une évaluation de ce lien de conformité entre les différentes caractéristiques de l'objet. Pour élaborer les dimensions constitutives de la qualité d'un dispositif, il faut donc caractériser l'objet et son système de référence. La procédure proposée consiste à repérer des référents dans le champ théorique à partir duquel l'évaluation se réalise, dans le contexte où elle se déroule et auprès des intervenants impliqués.

L'analyse de la documentation a montré que les auteurs s'entendent sur un certain nombre de composantes d'un programme ou dispositif (population ou besoins, objectifs, effets, etc.) bien que leur nombre et leur nature puissent varier d'un auteur à l'autre. Sur la base de cette documentation, sept composantes d'un dispositif ont été sélectionnées et définies: population, objectifs, ressources, environnement, actions, effets, espace/temps.

Population. Un dispositif se différencie du système auquel il appartient, parce qu'il est destiné à intervenir sur une population ayant des caractéristiques qui la distingue de la norme, réelle ou attendue. Cette population est définie en termes de lacunes, de manques ou de besoins. Cette population peut être constituée potentiellement de l'ensemble des personnes d'une société ou de certains groupes. Dans le cas des dispositifs étudiés, les troubles désignent une population et elles légitiment les dispositifs dont la forme scolaire s'écarte de celle de la classe régulière (Amigues, 2005).

Objectifs. Les objectifs sont, bien entendu, de natures diverses, bien que les objectifs pédagogiques prennent une place centrale pour l'école. Les objectifs sont de deux niveaux: ceux du dispositif et ceux du système auquel il appartient. Les objectifs du dispositif et les finalités du système se doivent d'être cohérents entre eux. Les finalités du système scolaire concernant les apprentissages scolaires, la socialisation ou l'équité sont ainsi complétées par des objectifs spécifiques au dispositif visant à les atteindre totalement ou partiellement.

Ressources. La mise en place d'un dispositif implique nécessairement, en plus d'une différenciation de la population et des objectifs, une différenciation des ressources. Il s'agit des ressources institutionnelles, des ressources humaines, des ressources financières, des ressources matérielles, des ressources temporelles, des ressources spatiales et des ressources scientifiques (Roegiers, 2003). Ces ressources peuvent être directes (mobilisation directe) ou indirectes (mobilisations conditionnelles) (De Ketele & Gerard, 2007). Ces ressources ne sont pas les mêmes que celles du système auquel il appartient sur les plans qualitatif, quantitatif et de leur organisation. Dans ce cadre, il faut distinguer entre les ressources prescrites, les ressources réelles et les ressources nécessaires ou souhaitables. L'organisation constitue les moyens prévus, « les manières de faire » pour atteindre l'objectif.

Environnement. L'environnement comprend les aspects physiques, légaux, financiers, sociaux, politiques du dispositif étudié, mais également des représentations sociales. L'environnement est perçu et perçoit. Par exemple, le contexte spécifique ou organisationnel dans lequel prend place le dispositif (par ex. : intégration/ségrégation) peut influencer potentiellement la perception du dispositif par différents acteurs sociaux (parents, enseignants, journalistes, etc.). Le cadre normatif constitue un élément essentiel du contexte d'un dispositif

(De Ketele & Gerard, 2007 ; Roegiers, 2003). Il est constitué de l'ensemble des prescriptions légales, éthiques, pratiques, déontologiques qui régissent, formellement ou informellement, directement ou indirectement, le dispositif.

Actions. Les actions que le dispositif entreprend visent tant à remplir sa mission stratégique prioritaire qu'à assurer sa fonction de survie (Foucault, 1977), c'est-à-dire à être suffisamment flexible pour se pérenniser. Les actions et rétroactions lient les composantes de base du dispositif et ses effets.

Effets. Les effets comprennent les résultats attendus du dispositif à court, moyen ou long terme, mais également les résultats inattendus de ce dernier. Les effets ont une influence sur la population, les ressources, les objectifs et l'environnement du dispositif. Les effets sont en fonction des composantes basiques du dispositif et de l'environnement et ils modifient, à leur tour, les composantes et l'environnement du dispositif. Ces effets sont donc à même de modifier le dispositif qui les a fait émerger.

Espace/temps. Un dispositif occupe un espace physique ou virtuel et il y évolue, au fil du temps, sous la pression de son environnement, de ses objectifs, de ses ressources, de ses actions et de ses effets. Étant opérationnalisées à des temps et des lieux différents, les « personifications » du dispositif sont amenées naturellement à « muter » tout en faisant évoluer et vieillir ce dernier.

Comme on le voit, les quatre premières composantes (populations, objectifs, ressources et environnement), soit les composantes basiques du dispositif, comportent différentes opportunités, mais également un certain nombre de contraintes. Par exemple, la population justifie tant le dispositif qu'elle en constitue une contrainte. Il en va de même avec les objectifs, l'environnement ou les ressources où un manque de disponibilité des ressources, ponctuel ou permanent constitue une contrainte tout comme le cadre légal et organisationnel dans lequel celles-ci sont déployées (Roegiers, 2003). Toutefois, c'est dans l'action et dans ses effets dans le temps et l'espace que peuvent être prise la mesure des contraintes réelles et des adaptations pour y faire face.

Proposition d'un modèle d'évaluation des dispositifs scolaires : les dimensions de la qualité

Sur la base de ces sept composantes, l'auteur a proposé la construction d'un modèle sur la base de liens de conformité entre composantes du dispositif, entre un référé et un référent. Ce travail a conduit à formaliser dix dimensions de la qualité, à savoir : la pertinence, l'adéquation, la congruité, la synergie, l'efficacité, l'efficience, la fiabilité, l'impact, le bien-fondé et la flexibilité. Ces dimensions sont bidirectionnelles et s'inscrivent dans un modèle (voir figure 1), donc nécessairement interreliées.

Pertinence. La pertinence d'un dispositif constitue le lien de conformité entre les objectifs du dispositif et la population pour laquelle il a été conçu. C'est la pertinence qui fournit la légitimité *a priori* du dispositif. Dans la première direction, l'objectif est le référé tandis que la population (et ses besoins) constitue le référentiel. En somme, les objectifs doivent répondre à des besoins clairement précisés dans la population (quels besoins ce programme vise-t-il à satisfaire? le programme répond-il aux besoins des participants?). La deuxième acception de la pertinence interroge le rapport de conformité entre la population (référé) et les objectifs (référentiel), c'est-à-dire si la population visée correspond à la population bénéficiaire. Cette question est particulièrement importante dans le cas de catégories mal définies et sujettes à interprétation (par ex. : troubles d'apprentissage, etc.). Dans le premier cas, il s'agit de se demander si l'on ne se trompe pas d'action tandis que dans le second, il y a lieu de se demander si on ne se trompe pas d'objet (Roegiers, 2003).

Adéquation. L'adéquation vise à évaluer le degré de conformité entre les objectifs/population du dispositif et les ressources prévues. Dans le premier cas, on interroge les ressources prévues (référé), c'est-à-dire leur suffisance en quantité, qualité et organisation en rapport aux objectifs/populations du dispositif (référentiel). On peut aussi se prononcer sur l'adéquation interne entre les ressources et leur organisation. Dans le second sens, il s'agit de se demander si l'objectif (référé) est atteignable avec les ressources qui sont attribuées (référentiel). Dans ce cas, l'ajustement possible ne touchera pas les ressources, mais plutôt l'objectif ou la population (par ex. : limitation de l'accès par manque de places).

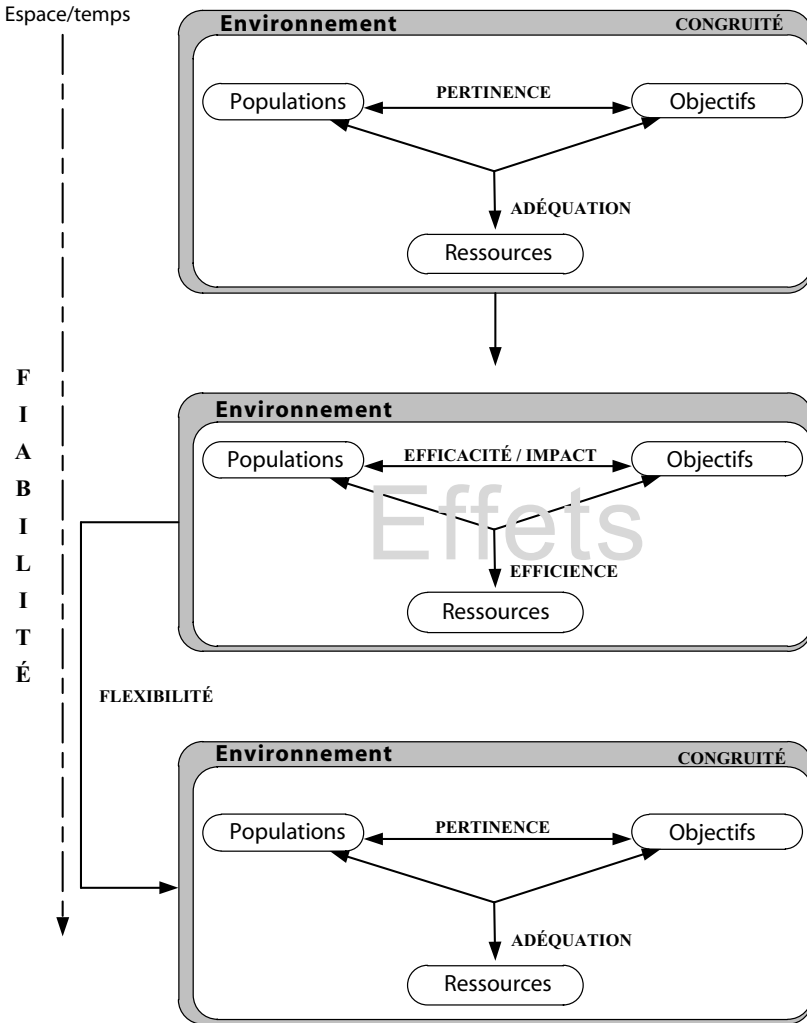


Figure 1. *Modèle d'évaluation de la qualité de dispositifs scolaires*

La première partie de la figure correspond au dispositif initial (temps zéro) avec ses composantes de base et son environnement. Des liens entre ces composantes, trois dimensions sont constituées. La deuxième partie correspond aux effets de l'action entreprise. Ces effets touchent directement les composantes de base. La troisième partie illustre le dispositif et son environnement par suite des effets directs de l'action et des effets de la rétroaction (temps 1). Entre le temps 0 et le temps 1, les composantes de base ont été modifiées. Cela influera, à son tour, sur les actions, les effets et la rétroaction, et ainsi de suite.

Congruïté. La congruïté concerne la prise en compte du contexte, de l'environnement, tant dans la mise en place du dispositif que dans sa pérennisation ou dans son fonctionnement. En somme, l'implantation d'un dispositif et son fonctionnement ne peuvent se faire sans tenir compte du système auquel il appartient, du contexte normatif, axiologique, politique, des représentations des acteurs externes. L'auteur peut définir la congruïté comme le lien entre les objectifs, la population et les ressources, d'une part, et son environnement, d'autre part. La première direction de la congruïté évalue le degré de conformité entre l'articulation des trois composantes basiques du dispositif : objectifs, population et ressources (référé) et son environnement (référentiel). Dans ce sens, il s'agit globalement de voir comment le dispositif entre ou non en conformité avec son environnement. Il s'agit, notamment, de l'adhésion de différents acteurs envers le dispositif. Dans la seconde direction, la congruïté évalue si l'environnement (référé) entre en conformité avec les composantes basiques du dispositif (référentiel). Il est possible, par exemple, de se demander en quoi l'environnement normatif et les acteurs externes participent à l'objectif d'intégration sociale ou professionnelle d'un dispositif destiné à des élèves à besoins spécifiques.

Ces trois premières dimensions sont issues des quatre composantes basiques du dispositif (population, objectifs, ressources et environnement) (voir figure 2).

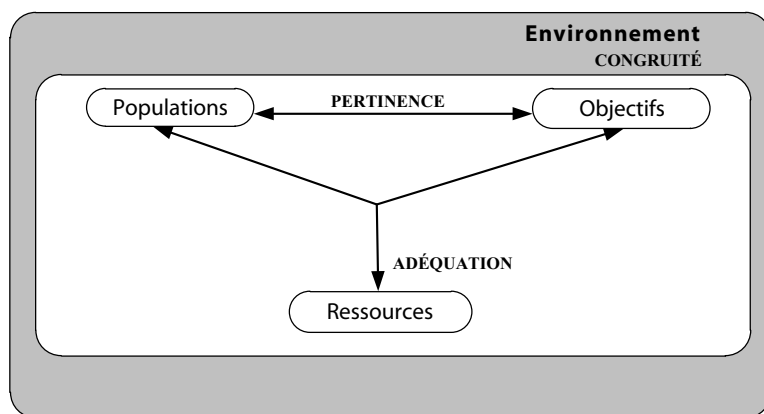


Figure 2. *Modèle d'évaluation de la qualité de dispositifs scolaires – Les composantes basiques*

Synergie. La synergie étudie les (inter) actions dirigées vers des objectifs, entre les ressources et une population. La synergie lie les composantes de base du dispositif et ses effets. Il s'agit tant de vérifier si les moyens (ressources et organisation) effectifs sont bien les moyens prévus que si le dispositif est mis en œuvre tel que planifié. L'analyse de la synergie combine l'identification des ressources qui sont réellement utilisées, les contraintes éprouvées et les stratégies réellement mises en œuvre. Par exemple, il est possible que certains professionnels ne soient peu, pas ou mal employés au sein d'un dispositif. Elle touche également les relations humaines et professionnelles. Cette dimension n'a qu'une seule direction, l'action, la direction miroir étant la flexibilité (la rétroaction), exercée, le plus souvent, par des intervenants différents (par ex. : gestionnaires, administrateurs, etc.).

Efficacité. L'efficacité est définie comme le lien de conformité entre les objectifs du dispositif et les effets obtenus (et prévus). Dans sa première acception, l'efficacité consiste globalement à vérifier si les objectifs fixés (référentiel) ont été atteints par les effets (référé) (par ex. : les effets constatés à court terme correspondent-ils aux objectifs initialement fixés ? les effets relevés chez les participants du dispositif se comparent-ils avec la situation des non-participants ?). Dans la seconde, on s'enquiert, à la lumière des effets obtenus (référentiel), de vérifier si les objectifs (référé) sont atteignables, dans quelle ampleur, etc. Il s'agit de se prononcer, en quelque sorte, sur la pertinence *a posteriori* des objectifs, mais également de la population. L'efficacité est directement dépendante des dimensions précédentes ; des manques sur les plans de la pertinence, de l'adéquation, de la congruité ou de la synergie influenceront directement l'efficacité du dispositif. Par exemple, une faible pertinence aura un impact certain, positivement ou négativement, sur l'efficacité. Des ressources insuffisantes, par exemple, risquent d'affecter tant les actions que les effets.

Efficience. L'efficience se définit comme le lien de conformité entre les ressources réellement mises en œuvre et les effets du dispositif. Cette dimension étudie le lien qui unit les ressources engagées et les effets prévus. Quatre types d'évaluation de l'efficience sont utilisés : minimalisation des coûts, coût-efficacité, coût-utilité, coût-bénéfice (Drummond, O'Brien, Stoddart, & Torrance, 1998). Dans la première direction, les ressources (référentiel) constituent le point de comparaison par rapport aux effets (référé). Cette évaluation interroge la suffisance des effets par rapport aux ressources déployées (minimalisation des coûts et coût-efficacité). Dans la seconde, les

effets servent de référentiel aux ressources (référé). Il s'agit de se demander, à la lumière des effets prévus, si cette dépense est utile ou bénéfique (par ex. : coût-utilité, coût-bénéfice).

Fiabilité. La fiabilité est constituée du lien de conformité entre les effets et l'espace/temps. La fiabilité du dispositif confronte l'efficacité avec les caractéristiques temporelles et spatiales du dispositif. Cette dimension est donc constituée de deux sous-dimensions: la durabilité et la régularité. La durabilité examine l'adéquation entre les effets et leur maintien dans le temps. Il faut la comprendre autant comme la mesure de l'évaluation du « vieillissement » du dispositif que de la permanence des effets. Par exemple, les besoins des participants ont-ils évolué depuis le début de la mise en œuvre en bénéficiant des effets du dispositif? La régularité, quant à elle, concerne la concordance entre les effets et leur maintien dans l'espace. La régularité consiste ainsi à vérifier si les résultats sont réguliers, dans la norme, pour chacune des personnifications du dispositif. La régularité est située à deux niveaux : les personnifications du dispositif et les bénéficiaires. La régularité se veut compréhensive. Elle cherche à comprendre et à expliquer pourquoi les résultats peuvent différer d'une personnification à l'autre ou d'un élève à l'autre, relevant pourtant du même dispositif.

Dans la première direction, l'espace/temps (référentiel) sert de point de comparaison aux effets (référé). On se demande si les effets se maintiennent dans le temps et dans l'espace. Par exemple, existe-t-il des différences d'effets entre les écoles, selon les époques? On vérifie, à la lumière des résultats obtenus (référentiel), si l'espace/temps (référé) affecte ceux-ci. Par exemple, des facteurs comme l'essoufflement ou le vieillissement des professionnels, des contraintes locales, etc., peuvent influencer la production des résultats des dispositifs. La seconde direction est ici impossible.

Impact. L'impact est le lien de conformité entre les effets imprévus et les objectifs du dispositif. Il s'agit d'une dimension de la qualité qui est paradoxale, car elle mesure ce qui n'est pas prévu au départ, ce qui est inattendu, ce que l'on ne cherche souvent pas. L'impact peut être une situation nouvelle (ou qui apparaît comme telle aux yeux des acteurs) ou encore un agrégat d'effets cumulés. De plus, les effets d'un dispositif sont théoriquement illimités et en interactions multiples et constantes avec les différentes caractéristiques du dispositif.

Ces effets d'impact peuvent être autant considérés comme positifs que négatifs, à court ou à long terme. Ces effets s'exercent tant sur la population, les ressources et les objectifs que sur l'environnement du dispositif. Dans la première direction, les effets (référé) sont analysés à la lumière des objectifs (référentiel). On analyse la nature, la taille de ces effets et leur valence (positive ou négative) en rapport à l'objectif poursuivi. Dans la seconde direction, les objectifs (référé) sont analysés à la lumière des effets non prévus (référentiel). Il s'agit de se demander, comme pour l'efficacité, au vu des résultats obtenus, si on ne doit pas modifier les objectifs du dispositif pour limiter ou accentuer des effets.

Bien-fondé. Le bien-fondé constitue le lien de conformité entre les effets du dispositif et les finalités du système dans lequel il s'inscrit. Ces finalités concernent les attentes du système et de la société dont il est issu et avec lequel il est en contact. Dans le cas de nos systèmes scolaires occidentaux, outre le développement cognitif des apprenants et la socialisation, l'équité est devenue un point de référence. L'émergence de dispositif, par sa différenciation, est d'ailleurs intrinsèquement liée à cette recherche d'équité. L'équité permet d'asseoir la légitimité *a posteriori* du dispositif. Ainsi, dans la première direction, l'équité (référentiel) sert de point d'analyse aux effets obtenus. En quoi ces effets favorisent-ils l'équité entre élèves, et ce, tant à l'intérieur du dispositif qu'à l'extérieur? Dans la seconde direction, les effets obtenus (référentiel) permettent de remettre en question la faisabilité de ces finalités (référé).

Flexibilité. La flexibilité d'un dispositif correspond à sa capacité d'adaptation et de changement. La flexibilité lie les effets aux composantes de base du dispositif. La flexibilité explore tant la connaissance de possibilité d'amélioration que les capacités de correction et de survie que possède un dispositif. C'est la flexibilité, en donnant des rétroactions, qui permet d'assurer la pérennité du dispositif. Les effets du dispositif (référentiel) questionnent les composantes de base (référé). La flexibilité cherche, premièrement, s'il y a des effets notables à prendre en compte et s'il faut apporter des correctifs. Une deuxième étape correspond aux actions entreprises pour limiter ou accentuer les effets et à l'évaluation de celles-ci.

Discussion

L'objectif de cet article était de proposer un modèle théorique permettant l'évaluation de deux dispositifs scolaires comparables, mais concurrents. Pour élaborer ce modèle, l'auteur a utilisé différentes sources: les sciences sociales et la recherche, les évaluations antérieures, les attentes des gestionnaires et praticiens, les connaissances de l'évaluateur et son expérience avec des dispositifs du même type et la pensée logique (Birckmayer & Weiss, 2000). Ce processus d'élaboration a impliqué la participation de trois groupes d'intervenants (macro, mezzo et micro).

Utilisant la référentialisation (Figari, 1994, 2011, 2006), l'auteur a proposé un modèle multidimensionnel de la qualité. Ce modèle utilise les caractéristiques de l'objet étudié (Fitzpatrick et al., 2011 ; Torres, 1996) en élaborant des rapports de conformité entre un référent et un référé constitués des composantes d'un dispositif (population, objectifs, ressources, actions, effets, espace/temps). Les dimensions sont bidirectionnelles et s'inscrivent dans un modèle holistique. Le modèle combine différents types d'évaluation: pertinence, processus, efficacité, efficience, impact et équité (Riddle & Dagenais, 2009). Ce modèle, faisant l'objet d'une concertation entre évaluateurs et intervenants, a été utilisé dans l'évaluation comparée de la qualité de deux dispositifs scolaires (inclusion et enseignement spécialisé) en structurant la sélection de 31 indicateurs et de 10 sous-indicateurs, utilisant une approche méthodologique mixte (prétests et post-tests, entretiens, sociogrammes, observations, etc.).

En plus des liens entre les composantes du dispositif, des liens entre les dimensions sont également présents. À un premier niveau, les actions (synergie et flexibilité) lient les composantes de base aux effets du dispositif. Ensuite, les composantes de base affectent les effets et inversement. Par exemple, un manque d'adéquation (par ex.: personnel pas ou mal formé) aura un effet probable sur l'efficacité ou l'impact du dispositif. Il en va de même avec la pertinence ou la congruité. Par ailleurs, les effets dus à l'action du dispositif portent tant sur les objectifs (par ex.: bien-fondé) que sur les ressources (par ex.: efficience) ou sur la population (par ex.: efficacité, impact, fiabilité). Ces effets directs et les rétroactions apportées par la flexibilité affectent ainsi à leur tour ces composantes de base. Le dispositif est ainsi amené à évoluer et vieillir dans l'espace et le temps. La population n'est plus exactement la même

(par ex. : amélioration des compétences de départ, déperdition, etc.), tout comme les ressources (par ex. : expérience, roulement du personnel, et autres ressources), etc.

Le modèle présenté s'inspire largement de deux modèles utilisés dans des phases antérieures de l'évaluation des dispositifs (Bouchard & Plante, 2002 ; De Ketele & Gerard, 2007). Il ne s'agit pas, à proprement parler, de création *ex nihilo* d'un modèle nouveau, mais de l'utilisation de modèles existants pour en apporter différentes améliorations sur la base d'une analyse critique et d'un processus de concertation avec les différents intervenants. Il s'en distingue sur quatre éléments fondamentaux : sa parcimonie, la bidirectionnalité des dimensions, l'intégration du bien-fondé et l'inclusion des effets combinés et différenciés (espace/temps).

La généralisation d'un tel modèle peut cependant être difficile, car les liens causals entre les différents éléments du dispositif restent supposés et demandent une validation pour mesurer la force de l'association entre les dimensions. Il reste nécessaire de le vérifier empiriquement. Dans le cadre d'évaluation des deux dispositifs scolaires en Communauté française de Belgique, ce modèle a montré son utilité pour appréhender une évaluation complète des dispositifs utilisant le concept de qualité. Son utilisation dans l'élaboration et l'organisation des indicateurs est apparue pertinente. Toutefois, ce modèle doit encore se confronter à d'autres réalités et échelles d'analyse pour démontrer son utilité et sa transférabilité.

Cependant, l'utilisation d'approches méthodologiques mixtes limite cette quête de validité objective et met en évidence le caractère subjectif de la validité (Cronbach, 1983). Levin-Rozalis (2009) rappelle que la validité, dans le cadre d'une évaluation, est principalement interne. C'est par le consensus qu'il recueille que le modèle se révèle valide. De plus, la validité est indirecte, car il s'agit d'un construit théorique. Dans le présent modèle, chacune des composantes du programme ou du dispositif est associée en mettant en évidence les relations théoriques entre le concept étudié (qualité) et ses dimensions. La validité est également jugée sur base de son exactitude quant au contenu sur la base d'avis d'experts qui évaluent si le modèle couvre toutes les facettes du concept et si les dimensions reflètent le concept de manière représentative. Enfin, dans une recherche d'une validité apparente, l'auteur s'est intéressé à l'avis de différents intervenants permettant d'indiquer le degré selon lequel on a l'impression qu'un modèle est valide et qu'il semble évaluer ce qu'il est censé évaluer.

Ce modèle prend en compte différentes limites attribuées habituellement aux modèles logiques. Premièrement, il tient compte des effets inattendus (impact) et de l'équité (bien-fondé). Ensuite, il se concentre sur plusieurs questionnements et non seulement sur ceux prévus par les gestionnaires du programme. Enfin, il inclut les effets de la rétroaction. Ce modèle possède une cohérence interne, c'est-à-dire une logique qui lui est propre, et une cohérence externe. De plus, l'approche comparative choisie, permettant de mieux discerner la qualité, intégrait la qualité mesurée et la qualité ressentie (Stake & Schwandt, 2006) tout en exigeant souplesse et neutralité. Par ailleurs, cette approche comparative pouvait prémunir contre des questions ou des réponses descriptives plutôt qu'évaluatives (Coryn et al., 2011).

Les limites tant de notre approche que de la théorie des modèles sont nombreuses et connues (Fitzpatrick et al., 2011). L'une de celles-ci, la sursimplification inhérente à ce type d'approche (Coryn et al., 2011 ; Fournier, 1995), peut aisément s'appliquer à notre modèle. Souhaitant un consensus large sur le modèle utilisé et adoptant une approche comparative, cela s'est fait naturellement au détriment de certaines spécificités ou simplement de l'établissement de standards. Bien que l'on s'appuie sur la participation et la concertation de différents groupes d'intervenants, le présent modèle se révèle plus théorique que pratique. Ce modèle reste abstrait, loin des réalités quotidiennes des intervenants. De plus, la participation et l'adhésion de différents intervenants sont ici une condition essentielle. Toutefois, bien qu'il ait fait l'objet d'un consensus sur les demandes de chaque niveau d'intervenants, l'auteur ne peut s'avancer sur la force de l'adhésion des intervenants à celui-ci à moyen et long termes ; différentes positions ont dû être rapprochées. En somme, le modèle se devait de pouvoir apporter des réponses partielles, propres à chaque groupe, mais également de permettre aux acteurs des différents niveaux d'adopter le regard des autres intervenants et une vision globale de l'évaluation des dispositifs.

Le modèle constitue ainsi un cadre initial de compréhension et de questionnements sur différentes dimensions de la qualité des dispositifs. Il a permis de faire évoluer, auprès de l'ensemble des intervenants, les intentions du début. L'auteur juge également que celui-ci offre la possibilité à différents acteurs (évaluateurs, acteurs institutionnels, enseignants, etc.) d'adopter une perspective holistique des dispositifs étudiés et d'ainsi mieux gérer leur complexité pour cerner les processus en œuvre et les correctifs éventuels à y apporter. Il représente un espace de discussion partagé. Ce modèle permet ainsi de

structurer l'information donnée aux intervenants pour qu'ils puissent procéder, à la lumière de celle-ci et prenant en compte le caractère systémique des dispositifs et de leur qualité, à des actions visant à améliorer la qualité des dispositifs. Cette approche souligne, en définitive, l'importance d'un cadre (ou métacadre) commun d'analyse permettant de prolonger et de réinvestir les acquis de l'évaluation vers la flexibilité des dispositifs étudiés. Ce cadre risque toutefois d'avoir une validité temporaire, car les intervenants pourront approfondir leur analyse et le modèle peut connaître des variations dans l'espace/temps. Cela ouvre la porte à des modèles multiples, différenciés selon les intervenants et le type de dispositifs (Hansen & Vedung, 2010) tout en maintenant un cadre commun ou de possibles interconnexions.

L'implication de l'évaluateur, partie prenante de ce processus, est également importante à relever. L'évaluateur a été associé à l'élaboration du dispositif inclusif et à son évaluation, ce qui crée un conflit d'intérêts potentiel (Stuffelbeam & Shinkfield, 2007). D'une part, cette implication a permis de s'assurer d'une connaissance approfondie du dispositif et des intervenants, ce qui est une condition essentielle à l'élaboration d'un modèle théorique. Cependant, cela suppose, d'autre part, que l'évaluateur risque, en étant associé à l'élaboration et à l'évaluation, de subir des influences et d'orienter ces recherches vers certains aspects plutôt que d'autres, de biaiser ses conclusions. Parallèlement, l'évaluateur n'est pas exempt d'un positionnement influencé par des valeurs. De plus, l'évaluation de dispositifs est une activité éminemment politique chargée de différentes intentions tant internes qu'externes. Malgré les efforts pour rationaliser le processus de la part des évaluateurs, l'évaluation de dispositifs n'est ainsi généralement pas perçue comme une entreprise neutre, mais comme un processus où des intérêts divergeant de groupes s'affrontent: l'un espérant des gains et l'autre redoutant des pertes. Cette situation a influencé les orientations vers ce type de modèles plus larges, neutres et consensuels, reconnaissant que les valeurs de l'ensemble des intervenants, évaluateurs y compris, et le processus d'objectivation et de contextualisation de celles-ci sont des aspects centraux et critiques de l'évaluation.

NOTES

1. Dans le cadre de cet article, les termes «dispositif» et «programme» sont utilisés de manière indifférenciée.
2. En Communauté française de Belgique, un enseignement spécialisé, distinct de l'enseignement ordinaire, est organisé pour les élèves ayant des troubles d'apprentissage (enseignement spécialisé de type 8). La population est définie par le décret de 2004 sur l'enseignement spécialisé; les deux dispositifs étant destinés aux élèves *«tout en ne manifestant pas de troubles de l'intelligence, de l'audition ou de la vision, ils présentent des troubles qui se traduisent par des difficultés dans le développement du langage ou de la parole et/ou dans l'apprentissage de la lecture, de l'écriture ou du calcul et dont la gravité est telle que, dans un premier temps, une intervention particulière dans le cadre de l'enseignement ordinaire ne peut suffire»* (Communauté française de Belgique, 2004, p. 7). Les élèves bénéficiant de ce type d'enseignement (spécialisé ou inclusif) doivent recevoir une attestation d'orientation vers l'enseignement spécialisé délivré par un organisme neutre (Centre psycho-médico-social - CPMS) à la suite d'un diagnostic et d'un protocole justificatif en cinq parties (médical, social, pédagogique/logopédique/orthophonique, psychologique et conclusion). Les parents doivent également accepter l'orientation proposée. Les élèves en inclusion devaient se soumettre au même processus et recevoir le même type de documents.

RÉFÉRENCES

- Adams, D. (1993). *Defining educational quality. Improving Educational Quality Project Biennial Report*. Arlington, VA: Institute for International Research.
- Amigues, R. (2005). Les dispositifs d'aide aux élèves en difficulté comme révélateurs de l'activité enseignante. In L. Talbot (Ed.), *Pratiques d'enseignement et difficultés d'apprentissage* (pp. 105-116). Toulouse: ERES.
- Arcaro, J. S. (1995). *Quality in education: An implementation handbook*. Delray Beach, FL: St. Lucie Press.
- Beamish, W. (2004). *Consensus about program quality: An Australian study in early childhood special education*. Brisbane, Australie: Griffith University.
- Behrens, M. (2007). Introduction. In M. Behrens (Ed.), *La qualité en éducation: pour réfléchir à la formation de demain* (pp. 3-18). Québec: Presses de l'Université du Québec.
- Birckmayer, J. D., & Weiss, C. H. (2000). Theory-based evaluation practice: What do we learn? *Evaluation Review*, 24, 407-431. doi:10.1177/0193841X0002400404
- Bouchard, B., & Plante, J. (2002). La qualité: mieux la définir pour mieux la mesurer. *Cahier du service de pédagogie expérimentale*, 11-12, 219-236.
- Campbell, C., & Rozsnyai, C. (2002). *Quality assurance and the development of course programmes*. Papers on Higher Education, Regional University Network on Governance and Management of Higher Education in South East Europe. Bucarest: UNESCO.
- Chen, H.T. (1990). *Theory-driven evaluation*. First edition. CA: SAGE.
- Chen, H. T. (2005). *Theory-driven evaluation*. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 415-419). Thousand Oaks, CA: SAGE.
- Communauté française de Belgique (2004). *Décret organisant l'enseignement spécialisé*. Communauté française de Belgique, Moniteur belge 03-06-2004.
- Communauté française de Belgique (2007). *Les indicateurs de l'enseignement*. ETNIC – Service des Statistiques de la Communauté française.
- Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, 32(2), 199-226. doi:10.1177/1098214010389321
- Council for Higher Education Accreditation (CHEA) (2002). *Glossary of Key Terms in Quality Assurance and Accreditation*. Retrieved from http://www.chea.org/international/inter_glossary01.html, last updated 23 October 2002.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed). Thousand Oaks, CA: SAGE.
- Cronbach, L. J. (1983). *Designing evaluation of educational and social programs*. San Francisco, CA: Jossey-Bass.
- Davidson, E. J. (2000). Ascertaining causality in theory-based evaluation. In P. J. Rogers, T. A. Hasci, A. Petrosino, & T. A. Huebner (Eds), *Program theory in evaluation: Challenges and opportunities* (pp. 17-26). San Francisco, CA: Jossey-Bass.

- Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.
- De Ketele, J.-M., & Gerard, F.-M. (2007). La qualité et le pilotage du système éducatif. In M. Behrens (Ed.), *La qualité en éducation: pour réfléchir à la formation de demain* (pp. 19-38). Québec: Presses de l'Université du Québec.
- Delvaux, B. (2000). Enseignement spécial. In Ministère de la Communauté française (Ed.), *Tableau de bord de l'enseignement* (pp. 100-102). Bruxelles: Service général de l'informatique et des statistiques.
- Doherty-Delorme, D., & Shaker, E. (2001). *Missing pieces II. An alternative guide to Canadian post-secondary education. 2000/2001 provincial rankings: Where do the provinces stand on education?* Canadian Centre for Policy Alternatives. Retrieved from <http://www.policyalternatives.ca/publications/pub4.html>.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Donaldson, S. I., & Lipsey, M. W. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The handbook of evaluation: Policies, programs, and practices* (pp. 56-75). London, UK: Sage.
- Drummond, M. F., O'Brien, B., Stoddart J. L., & Torrance, G. W. (1998). *Méthodes d'évaluation économique des programmes de santé* (2^e éd.). Paris: Economica.
- Figari, G. (1994, réédité en 2008). *Évaluer: quel référentiel?* Bruxelles: De Boeck Université.
- Figari, G. (2001). Us et abus de la notion de référentiel. In G. Figari & M. Achouche (Eds.), *L'activité évaluative réinterrogée: regards scolaires et socioprofessionnels* (pp. 310-314). Bruxelles: De Boeck.
- Figari, G. (2006). Les référentiels entre théorie et méthodologie. In G. Figari & L. Mottier Lopez (Eds.), *Recherche sur l'évaluation en éducation* (pp. 101-108). Paris: Harmattan.
- Figari, G. (2008). L'évaluation des dispositifs éducatifs. *Mesure et évaluation en éducation*, 31(3), 77-93.
- Figari, G. (2011). Us et abus de la notion de référentiel. In G. Figari & M. Achouche (Eds.), *L'activité évaluative réinterrogée: regards scolaires et professionnels* (pp. 310-314). Bruxelles: DeBoeck.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation. A lternative approaches and practical guidelines* (4th ed.). Boston, MA: Pearson.
- Foucault, M. (1977). *Le jeu de Michel Foucault. Dits et écrits (tome III)*. Paris: Gallimard.
- Fournier, D. M. (1995). Establishing evaluative conclusions: A distinction between general and working logic. In D. M. Fournier (Ed.), *New Directions in Evaluation. Reasoning in evaluative: Inferential links and leaps*. (Vol 68, pp. 15-32). San Francisco, CA: Jossey-Bass.
- Franssen, A., & Van Campenhoudt, L. (2004). *La consultation des personnels pédagogique, éducatif, paramédical, psychologique et social de l'enseignement spécialisé*. Bruxelles: Centre d'études sociologiques des Facultés universitaires Saint-Louis, Administration générale de l'Enseignement et de la Recherche scientifique.
- Greene, J. C. (2006). Evaluation, democracy, and social change. In J. Shaw, J. C. Greene, & M. Marks (Eds.), *Handbook of evaluation* (pp. 118-160). Thousand Oaks, CA: Sage.

- Hansen, M. B., & Vedung, E. (2010). Theory-based stakeholder evaluation. *American Journal of Evaluation*, 31(3), 295-313. doi:10.1177/1098214010366174
- Harvey, L. (1995). Editorial: The quality agendas. *Quality in Higher Education*, 1(1), 5-12.
- Hurteau, M., Valois, P., & Bossiroy, A. (2011). Jugement crédible en évaluation de programme: définition et conditions requises. *The Canadian Journal of Program Evaluation*, 25(2), 83-101.
- Juran, J. (1993). *Quality planning and analysis*. Maidenhead, UK: McGraw-Hill.
- Krueger, R. A. & Casey, M. A. (2000). *Focus groups: a practical guide for applied research*. Thousand Oaks, CA: Sage.
- Levin-Rozalis, M. (2009). Recherche et évaluation de programme. In V. Riddle & C. Dagenais (Eds), *Approches et pratiques en évaluation de programme* (pp. 31-49). Montréal: Les Presses de l'Université de Montréal.
- Lipsey, M. W., Rossi, P. H., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.
- Morgan, D. L. (1997). *Focus groups as qualitative research*. Thousand Oaks, CA: Sage.
- Otis-Wilborn, A., Winn, J., Ford, A., & Keyes, M. (2000). Standards, benchmarks, and indicators: Designing a framework for professional development of preservice and practicing teachers. *Teaching Exceptional Children*, 32(5), 20-28.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Pluye, P., Nadeau, L., Gagnon, M.-P., Grad, R., Johnson-Lafleur, J., & Griffiths, F. (2009). Les méthodes mixtes. In V. Riddle & C. Dagenais (Eds), *Approches et pratiques en évaluation de programme* (pp. 121-141). Montréal: Les Presses de l'Université de Montréal.
- Poole, B. (2010). Quality, semantics and the two cultures. *Quality Assurance in Education*, 18(1), 6-18. doi:10.1108/09684881011015963
- Porteous, N. L. (2009). La construction du modèle logique d'un programme. In V. Riddle & C. Dagenais (Eds), *Approches et pratiques en évaluation de programme* (pp. 85-105). Montréal: Les Presses de l'Université de Montréal.
- Riddle, V., & Dagenais, C. (2009). Introduction générale à l'évaluation de programme. In V. Riddle & C. Dagenais (Eds), *Approches et pratiques en évaluation de programme* (pp. 11-29). Montréal: Les Presses de l'Université de Montréal.
- Roegiers, X. (2003). *Analyser une action d'éducation ou de formation* (2nd ed.). Bruxelles: De Boeck.
- Rogers, P. J., Petrosino, A., Huebner, T. A., & Hacsí, T. A. (2000). Program theory evaluation: Practice, promise, and problems. In P. J. Rogers, T. A. Hacsí, A. Petrosino, & T. A. Huebner (Eds), *Program theory in evaluation: Challenges and opportunities: No. 87. New directions for evaluation* (pp. 5-14). San Francisco, CA: Jossey-Bass.
- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation. A systematic approach* (6th ed.). Thousand Oaks, CA: Sage.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Shaw, J. (2010). The diversity paradox: Does student diversity enhance or challenge excellence? *Journal of Further and Higher Education*, 33(4), 321-331.
- Stake, R., & Schwandt, T. (2006). On discerning quality in evaluation. In I. F. Shaw, J. C. Greene, & M. M. Mark (Eds), *Handbook of evaluation*. London: Sage.

- Stame, N. (2004). Theory-based evaluation and varieties of complexity. *Evaluation*, 10, 58-76. doi:10.1177/1356389004043135
- Stuffelbeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco, CA: Jossey-Bass.
- Torres, J.-C. (1996). Héxis et poïos: essai d'une analyse conceptuelle de la qualité. *Éducation permanente*, 126(1), 31-44.
- Tremblay, P. (2007). Évaluation de la validité et de l'efficacité interne de l'enseignement spécialisé primaire de type 8 en Wallonie. *Éducation – Formation – e*, 286, 9-21.
- Tremblay, P. (2011). Enquête longitudinale portant sur la transition postscolaire de jeunes ayant une déficience intellectuelle modérée à sévère. *Revue francophone de déficience intellectuelle*, 22, 86-97.
- UNESCO (2005). *Éducation pour tous – L'exigence de qualité*. Paris : UNESCO.
- Van Kemenade, E., Pupius, M., & Hardjono, T. W. (2008). More value to defining quality. *Quality in Higher Education*, 14(2), 175-185. doi:10.1080/13538320802278461
- Vienneau, R. (2002). Pédagogie de l'inclusion : fondements, définitions, défis et perspectives. *Éducation et francophonie*, XXX(2), 257-286.
- Vlăsceanu, L., Grünberg, L., & Pârlea, D., 2007, *Quality assurance and accreditation: A glossary of basic terms and definitions*. Édition révisée et mise à jour. Bucarest: UNESCO-CEPES.
- Wangenge-Ouma, G., & Langa, P. V. (2010). Universities and the mobilization of claims of excellence for competitive advantage. *The International Journal of Higher Education and Educational Planning*, 59(6), 749-764. doi:10.1007/s10734-009-9278-x
- Winch, C. (1996). *Quality and education*. Oxford, UK: Blackwell.

Date de réception : 31 mars 2011

Date de réception de la version finale : 28 mai 2012

Date d'acceptation : 18 juin 2012